Mathematical modeling and kinetic analysis of cellular signaling pathways

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Biophysik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von
Herr. MSc. Zhike Zi
geboren am 27.12.1979 in HUNAN, CHINA

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I
Prof. Dr. Lutz-Helmut Schön

Gutachter/innen: 1. Prof. Dr. Martin Vingron
2. Prof. Dr. Hermann-Georg Holzhütter
3. Prof. Dr. Edda Klipp

eingereicht: 23.04.2008

Datum der Promotion: 21.10.2008

Abstract

With growing interests in systems biology, mathematical models, paired with experiments, have been widely used for the studies on metabolic networks, gene regulatory networks and cellular signaling pathways. This dissertation employs the mathematical modeling and kinetic analysis method to study cellular signaling pathways, in particular, the receptor trafficking network and TGF-β signaling pathway. On the other hand, a systems biology markup language (SBML) based parameter estimation tool (SBML-PET), was developed for facilitating the modeling process.

A quantitative mathematical model is employed to investigate signal responses in different receptor trafficking networks by simultaneous perturbations of the ligand concentration and cell density. The computational analysis of the model revealed that receptor trafficking networks have potentially sigmoid responses to the ratio between ligand number and surface receptor number per cell, which is a key factor to control the signaling responses in receptor trafficking networks.

Using the SBML-PET software package, we proposed a constraint-based modeling method to build a comprehensive mathematical model for the Smad dependent TGF-β signaling pathway by fitting the experimental data and incorporating the qualitative constraints from the experimental analysis. Kinetic analysis results indicate that the signal response to TGF-β is regulated by the balance between clathrin dependent endocytosis and non-clathrin mediated endocytosis.

Keywords:
systems biology, signal transduction, mathematical model, TGF-beta

Zusammenfassung

Aufgrund des wachsenden Interesses an der Systembiologie werden zunehmend mathematische Modelle in Kombination mit Experimenten für die Analyse von Stoffwechselnetzwerken, Genregulationsnetzwerken und zellulären Signalweiterleitungswegen verwendet. Diese Dissertationsschrift benutzt die mathematische Modellierung und kinetische Untersuchungsmethoden zum Studium von zellulären Signalwegen, insbesondere des Netzwerkes zur Festlegung der Rezeptorlokalisation und des Tumorwachstumsfaktor-beta-Signalweges. Ergänzend wurde ein Computerwerkzeug (SBML-PET) entwickelt, das die Modellentwicklung unterstützt und der Parameterschätzung dient. Mit diesem Werkzeug kann man Modelle bearbeiten, die in der Systems Biology Markup Language (SBML) formuliert sind.

In dieser Arbeit wird ein quantitatives mathematisches Modell benutzt, um die Signalantwort in unterschiedlichen Rezeptorlokalisationsnetzwerken in Abhängigkeit von der Ligandenanzahl und der Zelldichte zu untersuchen. Die rechnergestützte Analyse des Modells hat ergeben, dass der Zustand eines Rezeptorlokalisationsnetzwerkes potenziell eine sigmoide Abhängigkeit von dem Verhältnis zwischen Ligandenanzahl und Oberflächenrezeptoranzahl pro Zelle zeigen. Dieses Verhältnis ist die entscheidende Kontrollgröße der Signalantwort in Rezeptorlokalisationsnetzwerken.

Mit Hilfe des SBML-PET Software-Paketes haben wir eine Modellierungsmethode mit Randbedingungen vorgeschlagen, um ein umfangreiches mathematisches Modell für den Smad-abhängigen TGF-β Signalweg zu erstellen und dessen Parameter aus experimentellen Daten unter Berücksichtigung qualitativer Nebenbedingungen zu fitten. Die Ergebnisse der kinetischen Untersuchung dieses Modells legen nahe, dass die Signalantwort auf einen TGF-β-Reiz durch die Balance zwischen clathrin-abhängier Endozytose und clathrin-unabhängiger Endozytose reguliert wird.

Schlagwörter:
Systembiologie, Signaltransduktion, Mathematische Modellierung, TGF-beta

Contents

List of Figures

List of Tables

Abbreviations

Co-Smad	common mediator Smad
EGF	epidermal growth factor
ES	evolution strategy
HaCaT	human keratinocyte cell line
I-Smad	inhibitory Smad
LHS	Latin hypercube sampling
LRC	ligand-receptor complex
LRi	internalized ligand-receptor complex
LRs	surface ligand-receptor complex
ODE	ordinary differential equation
P-Smad2	phosphorylated Smad2
RLRs ratio	ratio between ligand number and surface receptor number per cell
R-Smad	receptor regulated Smad
SBML	systems biology markup language
SBML-PET	systems biology markup language based parameter estimation tool
SD	standard deviation
Smad	Sma- and Mad-related protein
SRES	stochastic ranking evolution strategy
T1R	TGF-β type I receptor
T2R	TGF-β type II receptor
TGF-β	transforming growth factor β

Chapter 1 Introduction of Systems Biology

The community of systems biology has grown by leaps and bounds in the past a few years. Mathematical models, paired with experiments, have been widely used for the studies on metabolic networks, gene regulatory networks and cellular signaling pathways (Bentele et al., 2004; Di Ventura et al., 2006; Hoffmann et al., 2002; Klipp et al., 2005b; Schuetz et al., 2007). Although it is hard to achieve an agreement on its exact definition, systems biology could be "defined as the science that deciphers how biological functions arise from the interactions between components of living organisms" (Boogerd et al., 2007). There are two approaches of systems biology: top-down and bottom-up systems biology (Bruggeman and Westerhoff, 2007). The former focuses on X-omics researches involving the comprehensive measurement of experimental data in a large scale, for example, metabolomics, proteomics, or fluxomics and transcriptomics (Joyce and Palsson, 2006). In contrast, the latter starts with the interaction of the components of the system and employs mathematical modeling to gain insights about the mechanism and dynamics of the biological systems (Aldridge et al., 2006). This thesis uses the bottom-up modeling method to study the dynamic properties of cellular signaling pathways.

1.1 Why Do We Need Models?

Mathematical modeling of signaling pathways requires experimental data and prior knowledge about the pathways. Therefore, systems biology research could not be done without the contribution of traditional biochemical and molecular biology. However, classical biochemistry and molecular biology face the limitations such as inaccuracy, inability to deal with emergent properties and so on (Boogerd et al., 2007). The bottom-up systems biology approaches (kinetic modeling method) could fetch up these limitations by combining the experimental data and quantitative analysis.

Many sources of uncertainty including errors, noise, incomplete information and poor experimental design impose limitations on our confidence of the experimental data. The experimental data might lead to a misunderstanding of the true property of biological systems due to the limited number of experimental data. For example, we might get time course data

for protein X by a western blot experiment as shown in Figure 1.1A. The amount of protein X oscillates with the increase of time. The normal way for the experimentalist to judge the number of periods is to count the number of peaks appearing in the data by connecting the discrete data and the number of peaks is 3 (Figure 1.1B). However, modeling the system by incorporating the experimental data, we find the number of the oscillations during the investigated time might be 7, rather than 3, which is estimated by a simple analysis of the data (Figure 1.1C). Inspired by the model prediction, we could design a new experiment to test the predicted properties emerged from the model analysis.

The traditional biochemistry and molecular biology are useful to describe and discover the components of the biological system. However, most of the descriptions are carton-type hypotheses which give us the static and qualitative information about the system. The quantitative dynamic behavior of the system cannot be captured by such analyses. On the other hand, the nonlinear interactions of the components might exhibit some emergent properties such as bistability, oscillation and robustness, which are difficult to be found by experimental analysis but are easier to be investigated by kinetic analysis mathematical models.

1.2 Fundamentals of Modeling and Kinetic Analysis of Signaling Pathways

The mathematical form for modeling of signaling pathways depends on the properties of the studied system and the specific questions that are going to be answered. The most commonly used forms are ordinary and partial differential equation systems (ODE and PDE), which can be in either deterministic or stochastic format (Aldridge et al., 2006). Stochastic equations are useful to investigate the properties of the system arising from random fluctuations or noise. In this thesis, the most popular form, the deterministic ODE is used to study the properties of the signaling pathways. The ODE modeling approach can be applied to the modeling studies when the components of signaling pathway can be assumed to homogenously distribute in the cell and the stochastic effects can be ignored.

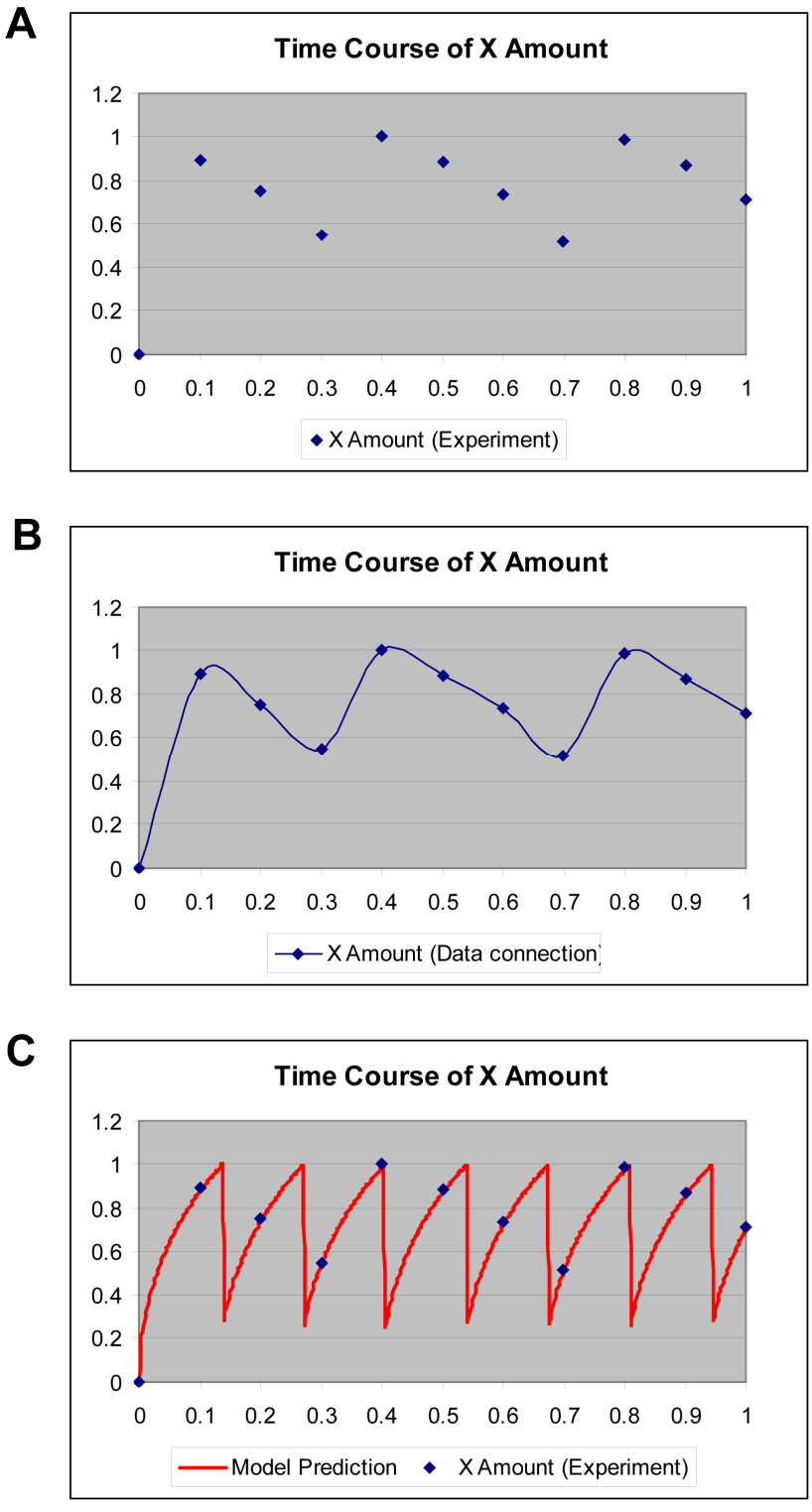

Figure 1.1 Illustration of the importance of model prediction. (A) Assumed experimental time course data for the amount of protein X. (B) The intuitive data connection predicts 3 oscillation periods. (C) Computational analysis of the underlying network suggests that protein X has oscillated 7 times.

1.2.1 Modeling of signaling pathways

The ODE modeling approach represents the concentration change of a component over time by an ordinary differential equation. For a certain component's concentration $[C_i]$ evolved over time, we calculate both the sum of the reaction rates producing C_i and the sum of the rates consuming C_i. The ordinary differential equation is determined by the subtraction of the former to the latter item as the following equation:

$$\frac{d[C_i]}{dt} = \sum v_{production} - \sum v_{consumption} \tag{1.1}$$

There are different ways to model the kinetics of reactions in signaling pathways. The commonly used kinetic methods are: law of mass action, Hill equation and Michaelis-Menten kinetics. Here, we use the example of ligand receptor interaction to illustrate how to use these kinetic approaches to describe the reactions steps.

The ligand and receptor interaction is a common process that happens in most of signaling pathways (Figure 1.2). With the binding of the ligand (L), the receptor (R) forms a ligand-receptor complex (LRC) with a characteristic kinetic constant k_{on}. On the other hand, the ligand-receptor complex can also dissociate to ligand and receptor with a kinetic constant k_{off}. The ratio of k_{off} to k_{on} is called dissociation constant K_d.

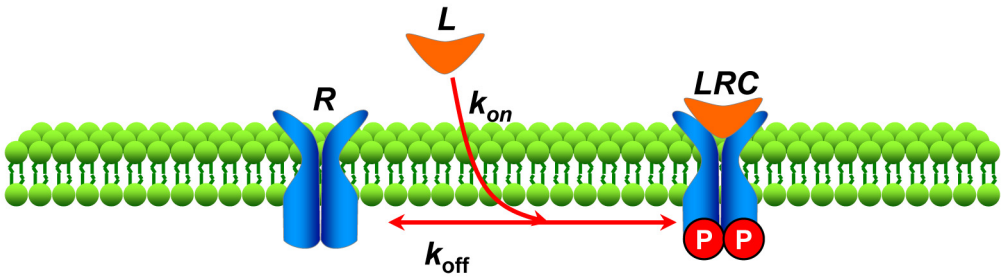

Figure 1.2 Ligand and receptor interaction

- **Law of mass action:** In 19th century, Guldberg and Waage introduced the law of mass action to describe the biochemical kinetics (Guldberg and Waage, 1879). The law of mass states that the reaction rate is proportional to the probability of the collision of the reactants. This probability is also proportional to the concentration of reactants to the power of their molecularity, the number of them entering the specific reaction (Klipp et

al., 2005a). The ligand receptor interaction scheme shown in Figure 1.2 can be simplified as the following reaction.

$$L + R \underset{k_{off}}{\overset{k_{on}}{\Leftrightarrow}} LRC \tag{1.2}$$

By the definition of mass action law, we can derive the concentration change over time of ligand, receptor and ligand receptor complex by the following ODE:

$$\frac{d[L]}{dt} = k_{off}[LRC] - k_{on}[L][R] \tag{1.3}$$

$$\frac{d[R]}{dt} = k_{off}[LRC] - k_{on}[L][R] \tag{1.4}$$

$$\frac{d[LRC]}{dt} = k_{on}[L][R] - k_{off}[LRC] \tag{1.5}$$

- **Hill equation:** Sometimes the receptor is not a monomer and it usually exists in a dimeric or tetrameric form with two or four identical binding sites to the ligand. The bound subunit has a cooperative effect on the later binding subunits, which means the affinity of the receptor to the later bound ligand are significantly increased by the already bound ligand. A typical example that has such a property is the binding of oxygen to hemoglobin (Hb). The early experimental research found that the fractional saturation of Hb with oxygen had a sigmoid response to the oxygen partial pressure. Hill explained the interaction between the oxygen binding sites of Hb subunits by the following fractional saturation (Hill, 1910):

$$Y = \frac{[L_n R]}{[R_t]} = \frac{[L]^n}{K_d + [L]^n} \tag{1.6}$$

where *n* is known as Hill coefficient.

In general, for a receptor with *n* subunits, the reaction rate of the receptor to the ligand can be approximated as:

$$v = V_{\max} Y = \frac{V_{\max}[L]^n}{K_d + [L]^n} \tag{1.7}$$

- **Michaelis-Menten kinetics:** When the signaling transduction reaction is catalyzed by an enzyme, the enzyme (kinase or phosphatase) is not consumed or produced by this reaction, but it may forms a temporary complex with the substance in the reaction. For such a reaction, we can use Michaelis-Menten kinetics to describe its reaction rate under the key assumption of quasi steady state approximation, which is valid when the enzyme concentration is much lower than the substrate concentration and when the enzyme is not allosteric. Michaelis-Menten kinetics is named for Leonor Michaelis and Maud Menten and has the following formulation:

$$v = \frac{V_{\max} S}{K_m + S} \tag{1.8}$$

where K_m is the Michaelis constant and is equal to the substance concentration that causes the half-maximal reaction rate V_{max}.

1.2.2 Model simulation

The description of the signaling pathways by an ODE mathematical model is the starting point to run computational simulations for the target pathways. In order to implement simulation for the ODE model *in silico*, we should know the initial conditions of the state variables (the concentration/amount of the molecules involved in the signaling pathways) and the parameter values (the rate constant values) that describe the speed of the reactions. Once the initial conditions and parameter values are known, we can implement the numerical simulation for the ODE by ODE solvers, for example, ODEPACK (Hindmarsh, 1983), Sundials (Hindmarsh et al., 2005) and so on.

The systems biology community has developed a Systems Biology Markup Language (SBML) to represent and exchange the models of biochemical reaction networks (Finney and Hucka, 2003; Hucka et al., 2003). SBML has a large international group of software developers and users. Some of the popular software packages for model construction, simulation and analysis are CellDesigner (Funahashi, 2003), COPASI (Hoops et al., 2006) and so on.

1.2.3 Model calibration (parameter estimation)

Since the initial conditions and parameter values are essential for the model simulation and analysis, we must know such information. However, most the initial conditions and parameter values are unknown or can not experimentally measured in practice. These values might be estimated by some way. The estimation of model parameters from experimental data remains a bottleneck for a major breakthrough in systems biology. The goal of parameter estimation for the ODE models is to find the possible parameters that minimize the difference between the computation simulation result and experimental analysis, which can formulized as:

$$\text{minimize } f_{obj} = \sum_{i=1}^{n} \sum_{t=t_0}^{t_{end}} \left(\frac{Y_{sim}(n,t) - Y_{data}(n,t)}{\sigma(n,t)} \right)^2 \tag{1.9}$$

where f_{obj} is called objective function, the cost function or sum of squared errors. Y_{sim} is the time course simulation result and Y_{data} is the corresponding experimental data. σ represents the noise or standard deviation of the experimental data.

The comparison of the simulation result to the experimental data that are used for parameter estimation is called “in-sample fit” (or “in-sample test”) and the parameter estimation process is regarded as “model calibration” (Aldridge et al., 2006). Different optimization algorithms can be used to implement parameter estimation work. The performance of the algorithms usually depends on the specific models (Moles et al., 2003; Zi and Klipp, 2006).

1.2.4 Model validation

Once we get the initial conditions and rate constants values, we can compare the model predictions with the experimental data. The “in-sample fit” comparison is necessary, but not enough for the model validation because it only give us the confidence on how good the model can reproduce the data that are used for parameter estimation. In order to evaluate the model’s prediction ability, we need to make some new predictions and compare it with new experimental data that are different from the data used for parameter estimation. This process is regarded as “out-sample fit” (or “out-sample” test) or model validation.

1.2.5 Steady state analysis

The concept of steady state is a mathematical idealization, which plays an important role in kinetic modeling (Heinrich and Schuster, 1996). A dynamic system is in steady state if the concentrations/amounts of the components do not change in time, which means the corresponding ordinary differential equations are zero. As an example of the protein X production and degradation system (Figure 1.3), we assume protein X is constitutively produced with a production rate constant V_r and degraded with a degradation rate constant k_t.

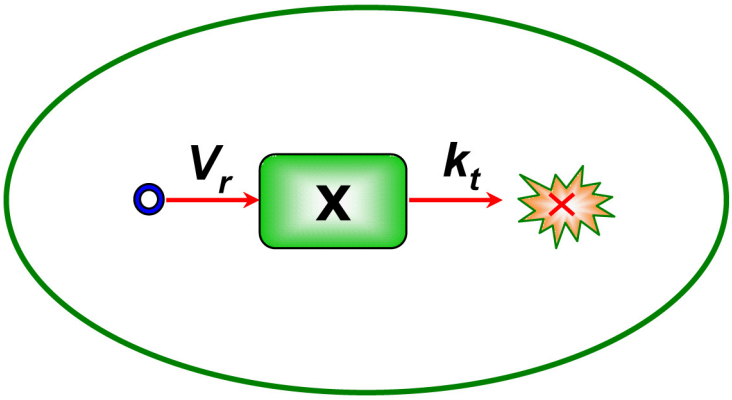

Figure 1.3 Production and degradation of protein *X*

We can describe the concentration change of protein X over time by the following ODE equation by assuming mass action kinetics for the degradation reaction:

$$\frac{d[X]}{dt} = V_r - k_t[X] \tag{1.10}$$

When there is no signal or other events affecting the behavior of X, the concentration of protein X will approach a steady state level after sufficient time. Then, it holds

$$\frac{d[X^{ss}]}{dt} = V_r - k_t[X^{ss}] = 0 \tag{1.11}$$

$$[X^{ss}] = \frac{V_r}{k_t} \tag{1.12}$$

Therefore, the steady state concentration of protein X is determined by the ratio of constant production rate V_r to the degradation rate constant k_t.

1.2.6 Sensitivity analysis

The mathematic models are valuable to reproduce what we get from the experimental analysis and predict the unobserved behaviors of the system with a set of parameters. However, many sources of uncertainty including noise of experimental data, absence of parameter information and poor understanding of the biological mechanism impose a limitation on our confidence in the model. Furthermore, intrinsic variability or noise of the system also affects the output of the model. Therefore, it is important not only to understand the dynamical properties of the model with particular parameter values, but also to further investigate the effect of their perturbations on the system (Zi et al., 2005). Sensitivity analysis is a powerful approach for investigating which parameters in a model have strong effects on the overall behavior. In addition to identifying key parameters in a model, sensitivity analysis is also valuable in pinpointing which parameters should be the focus of further experimental perturbation (Aldridge et al., 2006).

Sensitivity analysis has been widely utilized for the systems biology research (Bentele et al., 2004; Blower and Dowlatabadi, 1994; Cho et al., 2003; Feng et al., 2004; Hu and Yuan, 2006; Kholodenko et al., 2002; Mahdavi et al., 2007; Thomas et al., 1997; Zheng and Rundell, 2006; Zi et al., 2005). Two types of methods are developed for sensitivity analysis: local and global parameter estimation methods.

Local sensitivity analysis is the study of the changes in the model outputs with respect to small parameter variations, which are measured by the sensitivity coefficients. Mathematically, the sensitivity coefficients are the first order derivatives of a model output with respect to the model parameter:

$$S_{ij}^{unscaled} = \frac{\partial O_i}{\partial p_j} \tag{1.13}$$

where O_i is the i-th model output and p_j is the j-th parameter. $S_{ij}^{unscaled}$ is called unscaled or unnormalized sensitivity coefficient. When the model output and parameter is not zero, the

scaled or normalized sensitivity coefficients can be applied for the analysis, which is defined by the following formula:

$$S_{ij}^{scaled} = \frac{\partial O_i}{\partial p_j} \cdot \frac{p_j}{O_i} \tag{1.14}$$

Since local sensitivity analysis allows only a small perturbation of one parameter for each simulation, it investigates the sensitivity of the model output with respect to a particular point of the parameter set in the parameter space. However, there is probably not a single "true" point of parameter set that occurs in nature. It is likely that the parameters including rate constants and initial concentrations are varied in a large range depending on the specific cell types and cellular environments. For this reason, global sensitivity analysis algorithms are proposed to explore the possible non-linear effects of parameters by simultaneous variations of all the parameters with Monte-Carlo simulations (Bentele et al., 2004; Zheng and Rundell, 2006; Zi et al., 2005).

1.3 A Preview of What Will Come in This Dissertation

The following chapter of this thesis will first introduce the work on the regulation of signal response by cell density in receptor trafficking network (Zi and Klipp, 2007a). Computational analysis of the model revealed that receptor trafficking networks have a potentially sigmoid response on the ratio between ligand number and surface receptor number per cell, which is a key factor to control the signaling responses in receptor trafficking networks. In Chapter 3, a Systems Biology Markup Language (SBML) based Parameter Estimation Tool (SBML-PET) is going to be described. This tool is developed to estimate the parameter values in a model by fitting experimental data (Zi and Klipp, 2006). Since the estimation of model parameters from experimental data is very important for kinetic modeling and analysis of signal transduction pathways, SBML-PET will be employed in the following studies. The fourth chapter presents a specific modeling project, ie. the modeling of the Smad dependent TGF-β signaling pathway by constraint-based modeling method (Zi and Klipp, 2007b). The kinetic analysis results indicate that the signal response to TGF-β is regulated by the balance between clathrin dependent endocytosis and non-clathrin mediated endocytosis. Finally, the general discussion on the thesis gives an outlook for the future research.

Chapter 2 Cell Signaling Is Potentially Regulated by Cell Density in Receptor Trafficking Networks

2.1 Introduction

Cellular signaling processes, such as those associated with epidermal growth factor (EGF), transforming growth factor β, G-protein coupled receptor, Hedgehog and Notch pathways, elicits different cell type specific responses for cell proliferation, apoptosis, and differentiation (Fischer et al., 2006; Le Roy and Wrana, 2005; Seto et al., 2002; Tan et al., 2004). Cells communicate with their extracellular environments by the interaction between the receptors and ligand, which converts the information from the outside environment to inside cell responses such as cell proliferation, apoptosis, differentiation and growth. The receptors at cell surface can be internalized to early endosomes and also be recycled back, which is termed as receptor trafficking (Moore et al., 2007). Previous work performed for different systems has shown that receptor trafficking is a potential site for the control of signaling pathways (Childress et al., 2006; Di Guglielmo et al., 2003; Incardona et al., 2002; Schoeberl et al., 2002). In most biological experiments, the ligand concentration and cell density vary within a wide range among different systems. Traditionally, the extracellular medium compartment is ignored and theoretical models often assume that the ligand concentration in the medium is constant or follows a decaying function over time after the ligand addition (Lee et al., 2003; Schoeberl et al., 2002; Wiley and Cunningham, 1981). On the other hand, mathematical models of receptor trafficking networks usually ignored the trafficking of empty receptors or the recycling steps of receptors for the convenience of mathematical analysis (Kholodenko et al., 1999; Myers et al., 1987; Shankaran et al., 2007). However, there is less attention to systematically analyze whether these assumptions are valid and how much cellular signal responses are affected by various doses of ligand and cell density in the different systems.

In this work, we use quantitative mathematical models to investigate the cellular signal responses in different receptor trafficking networks by simultaneous variations of ligand concentration and cell density. From the analysis, it becomes apparent that the assumption, taken for granted in previous studies, that the ligand concentration in the medium is constant

or decaying function is only valid in some special conditions. Computational analysis of the model revealed that receptor activities have potentially sigmoid dependence on the ratio between ligand number and surface receptor number per cell, which is a key factor to control the signaling responses in receptor trafficking networks. Furthermore, cell density also affects the robustness of dose-response curve upon the variation of binding affinity. As a specific test of the model, we found that the model predictions are consistent with experimental observations of EGF dose-response curves.

2.2 Mathematical Models of Receptor Trafficking Networks

Figure 2.1 presents a schematic description of the receptor trafficking networks in this work. Several detailed kinetic models of receptor trafficking networks have been previously described (DeWitt et al., 2001; Knauer et al., 1984; Myers et al., 1987; Schoeberl et al., 2002; Starbuck and Lauffenburger, 1992; Wiley and Cunningham, 1981). Here, we revert to the canonical receptor trafficking models that are simple but include the basic features of the investigated systems. The canonical models either ignored the recycling step of receptors or didn't consider the trafficking of the "empty receptors" for the simplicity of mathematical analysis. Here we refer "empty receptors" to the receptors without the binding of ligand and "occupied receptors" to those bound by the ligand.

For the purpose of our general analysis, we considered eight possible topologies of receptors trafficking networks. The M1 receptor trafficking network has a general topology which includes the *de novo* appearance of surface receptor, ligand-receptor interaction, internalization, recycling and degradation of both empty and occupied receptors. The other 7 networks are derived from the general receptor trafficking network M1 by the possible knock-out mutations of corresponding recycling and dephosphorylation steps. Figure 2.1 gives an overview of eight topologies of the receptor trafficking networks.

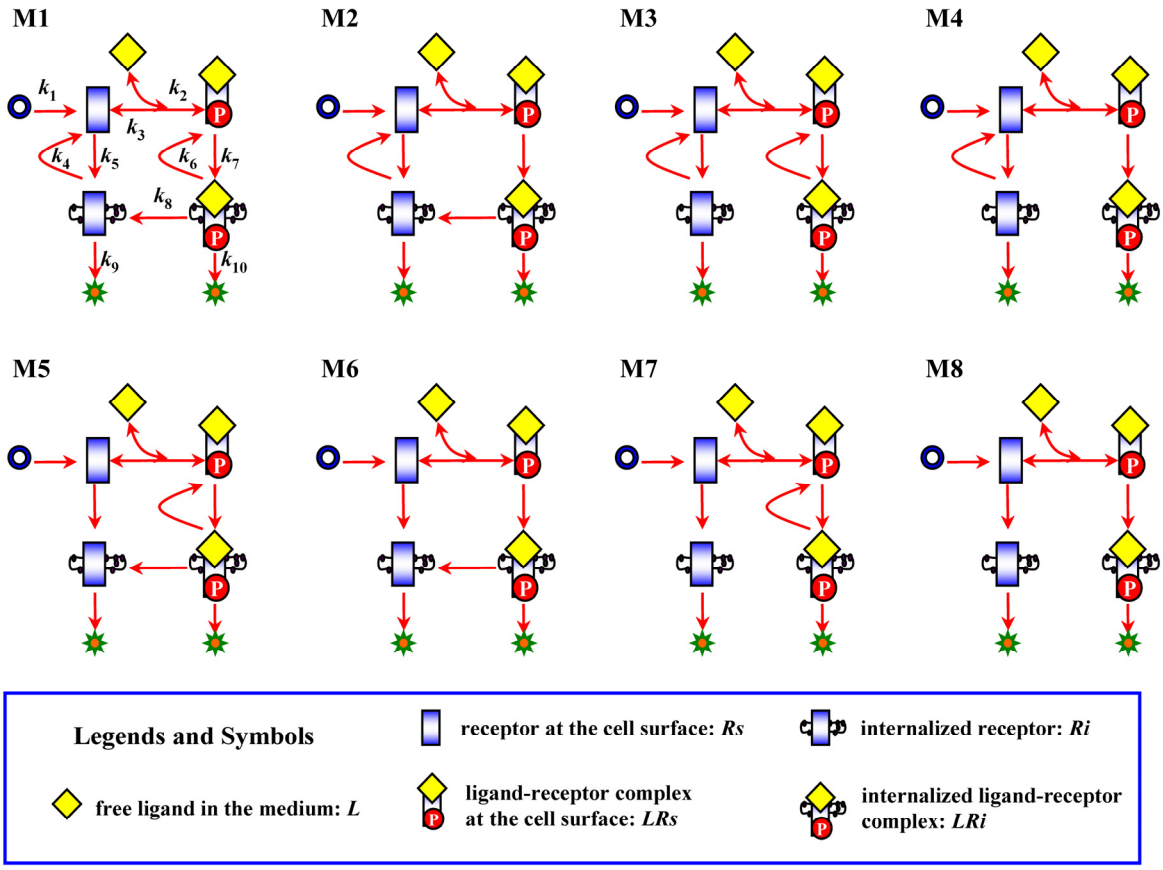

Figure 2.1 Schematic representation of the receptor trafficking networks

Similar to the previous canonical model, we model *de novo* appearance of surface receptor as a constant production rate k_1 without considering the possible regulation of receptor production. Extracellular ligand L can reversibly bind free surface receptors Rs with forward association rate constant k_2 and disassociation rate constant k_3 to form a receptor-ligand complex LRs at cell surface. The empty and occupied surface receptors are internalized with characteristic internalization rate constants k_5 and k_7, respectively. On the other hand, the empty and occupied receptors inside the cell can be recycled back to cell surface with recycling rate constant k_4 and k_6, respectively. The degradation rate constants of empty and occupied receptors are set as k_9 and k_{10}, respectively. We also considered the effect of phosphatases on the dephosphorylation of the activated receptors with rate constant k_8, which has not been taken into account in the previous models. Another important feature of our model is that we distinguished the extracellular medium and the cell as two different compartments for ligand in the medium and receptor trafficking molecules in the cell.

The rates of concentration change for the components in the general receptor trafficking network (M1 in Figure 2.1) can be written as:

$$\frac{d[Rs]}{dt} = -(k_2[L] + k_5)[Rs] + k_4[Ri] + k_3[LRs] + k_1 \tag{2.1}$$

$$\frac{d[Ri]}{dt} = k_5[Rs] - (k_4 + k_9)[Ri] + k_8[LRi] \tag{2.2}$$

$$\frac{d[LRs]}{dt} = k_2[L][Rs] - (k_3 + k_7)[LRs] + k_6[LRi] \tag{2.3}$$

$$\frac{d[LRi]}{dt} = k_7[LRs] - (k_6 + k_8 + k_{10})[LRi] \tag{2.4}$$

$$\frac{d[L]}{dt} = \frac{(-k_2[L][Rs] + k_3[LRs])V_{cell}}{V_{extra}} \tag{2.5}$$

where V_{cell} is the cell volume size and V_{extra} is the volume size of extracellular medium per cell. The value of V_{extra} is relative cell density and can be calculated as following equations:

$$V_{extra} = \frac{V_m}{N_c} = \frac{1\times10^{-3}}{D} - V_{cell} \tag{2.6}$$

$$D = \frac{1\times10^{-3}}{V_{cell} + V_{extra}} \tag{2.7}$$

where V_m is the total volume of the extracellular medium; N_c is the total number of the cells in the medium; D is the cell density in the culture, which is defined as how many cells per ml of media culture.

Most of the parameter values in the model of the receptor trafficking network have been experimentally measured and reported. In order to keep the consistency of the kinetic parameters, we choose the average reported values from the well-studied and well-documented EGF signaling pathway. Detailed information about the derivation of parameter values is described in next section.

For the other 7 mutated receptor trafficking networks (M2-M8 in Figure 2.1), they have the same ordinary differential equations as equations (2.1-2.5) with the exception that the rate

constants for the corresponding knock-out mutated steps are set zero. The initial conditions of the empty ligand-receptor complex *Rs* and *Ri* are set as the steady state values in the case where there is no extracellular ligand present, which are:

$$[Rs]_0 = \frac{k_1(k_4 + k_9)}{k_5 k_9} \tag{2.8}$$

$$[Ri]_0 = \frac{k_1}{k_9} \tag{2.9}$$

The initial conditions of the occupied ligand-receptor complex *LRs* and *LRi* are zero.

2.3 Materials and Methods

2.3.1 Derivation of parameter values and variation ranges

For most the parameters, including the binding constant, ligand concentration, cell densities, the biological relevant range spans several orders of magnitude. The rational which guided the choice of relative parameters is described in the following.

- **Cell Volume (V_{cell}):** Mammalian cell volumes vary among different cell types. An idealized cell could be considered as a sphere with a diameter of 15 μm, which results in a cell volume of 1.8×10^{-12} Liter. This typical value is approximate to the reported cell volume in mammalian cells (Conlon and Raff, 2003; Savini et al., 2000; Schoeberl et al., 2002; Watt et al., 1988).

- **Cell Density (D) and the extracellular medium volume per cell (V_{extra}):** For the cell culture system, a typical cell culture experiment would have a cell density of 2.5×10^{6} cells/ml. Thus, $V_{extra} = 4\times10^{-10}$ Liter/cell. In this study, we focus on the cell culture system and vary the extracellular medium volume per cell (V_{extra}) within the range of $1\times10^{-12} \sim 1\times10^{-8}$ Liter/cell. The corresponding range of cell density (D) in the culture is $1\times10^{5} \sim 3.6\times10^{8}$ cells/ml.

- **Ligand Concentration:** We choose the ligand concentration in the medium in the range of 0.001 ~ 1000 nM, which spans the reported ligand concentrations in the experiments.

For the other receptor trafficking parameters, there are many experimental measured values reported in different signaling pathways. In order to keep the consistency of the kinetic parameters, we choose the average reported values from the well-studied and well-documented EGF signaling pathway. The data from other signaling pathways such as TGF-β, LDL and TNF signaling pathways are usually within range of the reported data from the EGF signaling pathways. Therefore, the average of the data from EGF signaling pathway will not lost the generality of the application to other pathways.

- ***De novo* production rate constant of surface receptors (k_1):** This value is also symbolized as V_r in the literatures. The reported values of k_1 are shown in Table 2.1 and the average value of them is 0.45 nM/min.

Table 2.1 Receptor production rate constant (k_1) values

k_1 Values	Remarks/Notes	Reference
0.5 nM/min	hEGFR for B82 mouse fibroblasts	(DeWitt et al., 2001)
0.2 nM/min	for B82 mouse fibroblasts	(Starbuck and Lauffenburger, 1992)
0.58 nM/min	HF cells	(Knauer et al., 1984)
0.53 nM/min	HF cells	(Wiley and Cunningham, 1981)

- **Association and dissociation rate constant of ligand-receptor interaction (k_2 and k_3):** The association rate constant (k_2) of ligand-receptor binding spans several order of magnitudes. However, the variation of dissociation rate constant (k_3) of the ligand-receptor complex is small. On the other hand, previous work suggest that high-affinity binding results from an elevated association rate constant for high-affinity receptors as compared to that for low-affinity receptors, which means that high-affinity is controlled by the association rate constant (k_2) rather than the dissociation rate constant (Felder et al., 1992). Based on the reported data, the rate constant of the association rate constant (k_2) is varied within the range of 0.01 ~ 10 $(nM)^{-1}min^{-1}$ (1.0×10^7 ~ 1.0×10^{10} $M^{-1}min^{-1}$), which spans the reported data in Table 2.2. The typical value for k_2 is set as 1.28 $nM^{-1}min^{-1}$ (1.28×10^9 $M^{-1}min^{-1}$), which is the mean reported values. The dissociation rate constant of ligand-receptor complex is set as 0.2 min^{-1}, which is the average of the

following reported data shown in Table 2.3. Combining the rate constant k_2 and k_3, the value of binding affinities (K_D) can be varied within the range of 0.02 nM ~ 20 nM.

Table 2.2 Ligand receptor association rate constant (k_2) values

k_2 Values	Remarks	Reference
6.3×10^7 $M^{-1}min^{-1}$	hEGFR pH 7.4 for B82 mouse fibroblasts	(French et al., 1995)
6.2×10^7 $M^{-1}min^{-1}$	mEGFR pH 7.4 for B82 mouse fibroblasts	(French et al., 1995)
7.2×10^7 $M^{-1}min^{-1}$	for B82 mouse fibroblasts	(Starbuck and Lauffenburger, 1992)
1.8×10^8 $M^{-1}min^{-1}$	for mouse fibroblasts	(Knauer et al., 1984)
1.74×10^8 $M^{-1}min^{-1}$	for human fibroblasts	(Knauer et al., 1984)
3.72×10^7 $M^{-1}min^{-1}$	HeLa cell (high affinity)	(Berkers et al., 1991)
1.98×10^8 $M^{-1}min^{-1}$	HeLa cell (low affinity)	(Berkers et al., 1991)
5.9×10^9 $M^{-1}min^{-1}$	high affinity in HER14	(Felder et al., 1992)
1.5×10^7 $M^{-1}min^{-1}$	low affinity in HER14	(Felder et al., 1992)
7.2×10^9 $M^{-1}min^{-1}$	high affinity in A431	(Felder et al., 1992)
1.2×10^7 $M^{-1}min^{-1}$	low affinity in A431	(Felder et al., 1992)

Table 2.3 Ligand-receptor complex dissociation rate constant (k_3) values

k_3 Values	Remarks	Reference
0.26 min^{-1}	hEGF, pH 7.4 in B82 mouse fibroblasts	(French et al., 1995)
0.16 min^{-1}	mEGF, pH 7.4 in B82 mouse fibroblasts	(French et al., 1995)
0.228 min^{-1}	estimated, in HeLa cell	(Schoeberl et al., 2002)
0.34 min^{-1}	in B82 mouse fibroblasts	(Starbuck and Lauffenburger, 1992)
0.174 min^{-1}	in HER14 cell	(Felder et al., 1992)
0.1 min^{-1}	in A431 cell	(Felder et al., 1992)
0.12 min^{-1}	in CD533 cell	(Felder et al., 1992)

- **Internalization rate constant of empty and occupied receptors (k_5 and k_7):** The internalization rate constant of empty receptor is set as the average reported values

(Table 2.4), which is 0.13 min^{-1}. The internalization of the occupied receptor is set as the average reported values (Table 2.5), which is 0.19 min^{-1}.

Table 2.4 Empty receptor internalization rate constant (k_5) values for M1-M4

k_5 Values	Remarks	Reference
0.03 min^{-1}	for empty receptor, in HB2 cell	(Burke and Wiley, 1999)
0.11 min^{-1}	for empty receptor, in 48R cell	(Burke and Wiley, 1999)
0.12 min^{-1}	for empty receptor, in161 cell	(Burke and Wiley, 1999)
0.22 min^{-1}	for empty receptor, in 184 cell	(Burke and Wiley, 1999)
0.16 min^{-1}	for empty receptor, in 184A1 cell	(Burke and Wiley, 1999)

Table 2.5 Occupied receptor internalization rate constant (k_7) values

k_7 Values	Remarks	Reference
0.32 min^{-1}	for occupied receptor, in mouse fibroblasts	(Knauer et al., 1984)
0.14 min^{-1}	for occupied receptor, in human fibroblasts	(Knauer et al., 1984)
0.165 min^{-1}	for occupied receptor, in B82 fibrobalsts	(Tzafriri and Edelman, 2007)
0.15 min^{-1}	for occupied receptor, in HB2 cell	(Burke and Wiley, 1999)
0.2 min^{-1}	for occupied receptor, in 48R cell	(Burke and Wiley, 1999)
0.27 min^{-1}	for occupied receptor, in 161 cell	(Burke and Wiley, 1999)
0.17 min^{-1}	for occupied receptor, in 184 cell	(Burke and Wiley, 1999)
0.15 min^{-1}	for occupied receptor, in 184A1 cell	(Burke and Wiley, 1999)
0.149 min^{-1}	for occupied receptor, in neuroblastoma N2a cell	(Salazar and Gonzalez, 2002)

- **The recycling rate constant of empty and occupied receptor (k_4 and k_6):** The values of k_4 and k_6 are set as the average reported values of the recycling rate constant of the receptors inside of the cell (Table 2.6), which is 0.19 min^{-1}. For the receptor trafficking networks M5-M8 in Figure 2.1, the recycling step of empty receptor is mutated and the corresponding rate constant (k_4) is set to be zero. In these receptor trafficking networks, the corresponding internalization rate constant is turnover rate constant of the surface

receptors which is set as 0.001 min^{-1}. This typical value leads to the number of total receptor being 2.7×10^5 molecules/cell.

Table 2.6 Recycling rate constant (k_4 and k_6) values

k_4 and k_6 Values	Remarks	Reference
0.058 min^{-1}	receptor cycling rate constant, in B82 fibroblasts	(Starbuck and Lauffenburger, 1992)
0.2 min^{-1}	receptor cycling rate constant, in HB2 cell	(Burke and Wiley, 1999)
0.3 min^{-1}	receptor cycling rate constant, in 184A1 cell	(Burke and Wiley, 1999)

- **The dephosphorylation rate constant of active receptor complex (k_8):** The value of k_8 is set to be 0.1 min^{-1}, which is in the same magnitude as other parameters. Randomly choose a value within the range of 0.01~1 leads to a similar conclusions present in this study.

- The degradation rate constant of empty and occupied receptor ($\boldsymbol{k}_9$ and $\boldsymbol{k}_{10}$): The half-life time of the empty receptors are measured approximately 10 hours (Burke and Wiley, 1999; Schlessinger, 1988; Soderquist and Carpenter, 1986), which give us the typical degradation rate constant of empty receptor (k_9) as 0.001 min^{-1}. The half-life of occupied receptors is usually much shorter than the empty receptors, which is about 1 hour (Burke and Wiley, 1999; Schlessinger, 1988; Soderquist and Carpenter, 1986). The typical degradation rate constant of occupied receptor is estimated as 0.01 min^{-1}.

2.3.2 Measurement of signal response

Accumulating evidence indicates that downstream signal responses are coupled to the activated receptor complex at cell surface or in the endosome (Nagashima et al., 2007; Wiley, 2003). Therefore, we used the integrated response of the activated ligand-receptor complex at cell surface (*LRs*) and the internalized ligand-receptor complex (*LRi*) to evaluate the signal response of receptor trafficking pathway. We investigated the time course of *LRs* and *LRi* for 10 hours. The integrated response of *LRs* is equivalent to the following:

$$I_s = \int_0^t [LRs]dt \tag{2.10}$$

This expression can be evaluated approximately by the discrete summation:

$$\int_0^t [LRs]dt = \frac{\Delta t}{2}([LRs](t_0) + 2[LRs](t_1) + ... + 2[LRs](t_{n-1}) + [LRs](t_n)) \tag{2.11}$$

A similar expression is also applied to the integrated response of the internalized ligand-receptor complex (*LRi*).

2.3.3 Variations of ligand concentration, cell density and binding affinity

The ligand concentration, cell density and binding affinity of the same receptor trafficking network in the same type of cell are usually varied in different experimental conditions. For each analysis, we run 100~10000 simulations by using the Latin Hypercube Sampling (LHS) method (Mckay et al., 1979; Zi et al., 2005), which is an efficient method to sample random parameter vectors while guaranteeing that individual parameter ranges are evenly covered. Detailed information about the Latin Hypercube Sampling method is described in Appendix A. Further analyses indicate that the results are not sensitive to the number of simulations. The ranges of the corresponding parameters used in the samplings span several orders of magnitude to make sure that previously reported values are covered, which are: initial ligand concentration, 0.001~1000 nM; binding affinity, 0.02~20 nM; cell density 1×10^5 ~ 3.6×10^8 cells/ml.

2.4 Results

2.4.1 Cell signal response behaviors differently in low and high cell density cultures

In the models of this study, we distinguish the extracellular medium and the intracellular space as two different compartments for ligand in the medium and receptor trafficking molecules in the cell. Such a treatment is important for quantitatively investigating the dynamics of the signal response because ignoring the extracellular medium compartment will not correctly represent the dynamics of ligand and other receptor trafficking components. When cells are exposed to the same concentration of ligand in the medium with low cell density and high cell density, traditional mathematical models will give us the same

dynamical behaviors of the components for a single cell. However, this conclusion is not always true for the cells in the cultures with different cell density.

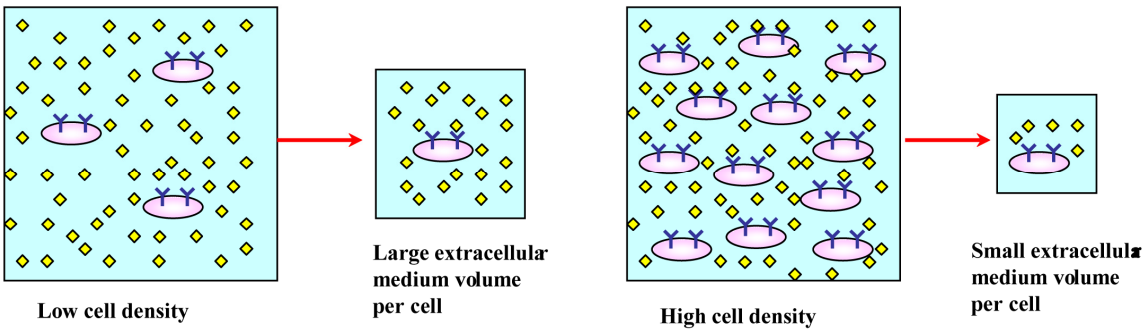

Figure 2.2 The effect of cell density on the volume of extracellular medium per cell

As illustrated in Figure 2.2, the cell in low cell density cultures has more available ligands per cell than that in high cell density cultures because the volume of extracellular medium per cell in low cell density cultures is larger than that in high density cultures. Therefore, the time course dynamics of the ligand concentration in the medium are different in the low cell density and high cell density cultures (Figure 2.3A). Furthermore, the signal response of surface and internalized ligand-receptor complex in low cell density cultures is stronger than that in high cell density cultures (Figure 2.3B-2.3C).

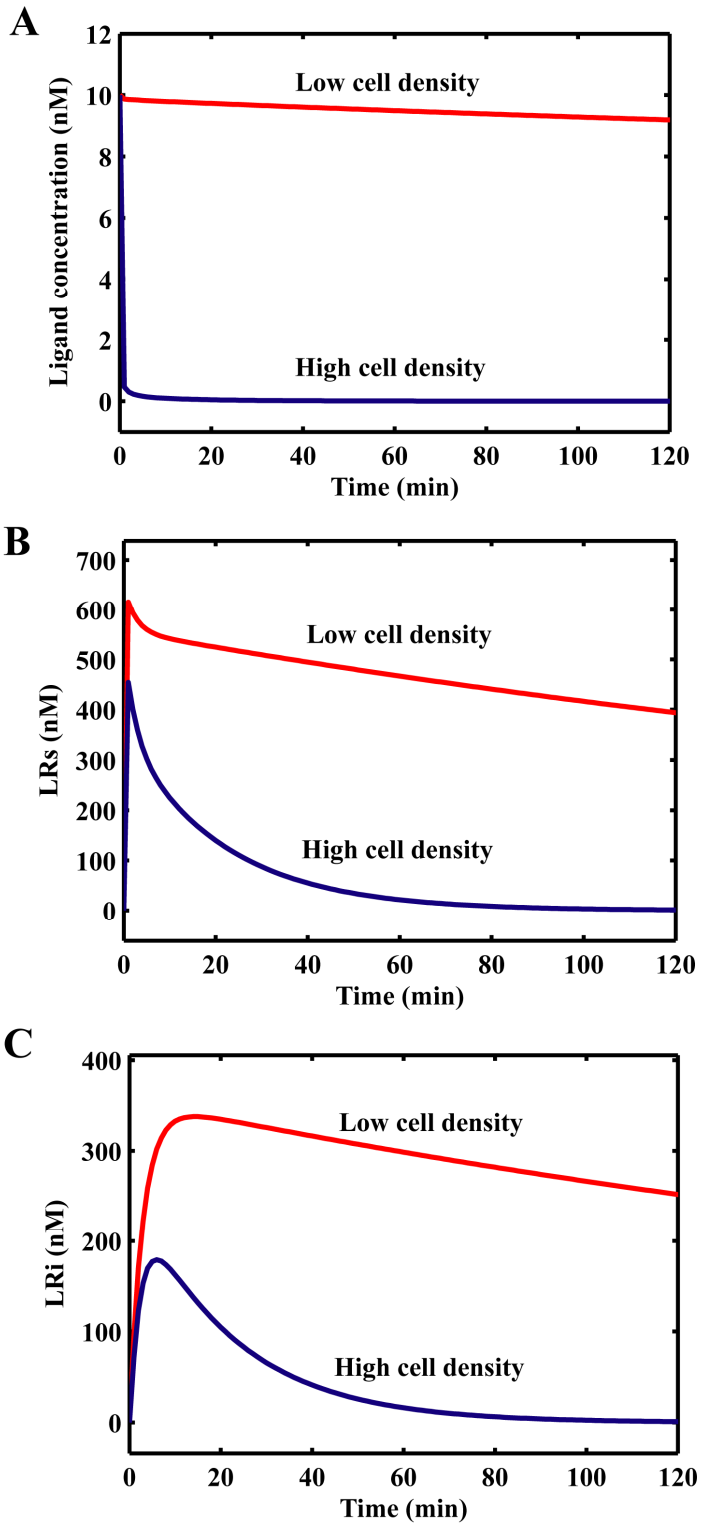

Figure 2.3 Effect of cell density on the signal response. (A) Time course of ligand concentration in the medium. (B) Time course of *LRs* in high and low cell density cultures stimulated with same dose of ligand. (C) Time course of *LRi* in high and low cell density cultures stimulated with same dose of ligand.

On the other hand, we computed the integrated response of the ligand-receptor complex at cell surface (*LRs*) with different ligand concentration for different cell density cultures. The computational analysis indicates the signal response of cell shows different dose-response curve in different cell density of cultures. The dose-response curve of the signal is shifted to the right when the cell density is increased (Figure 2.4).

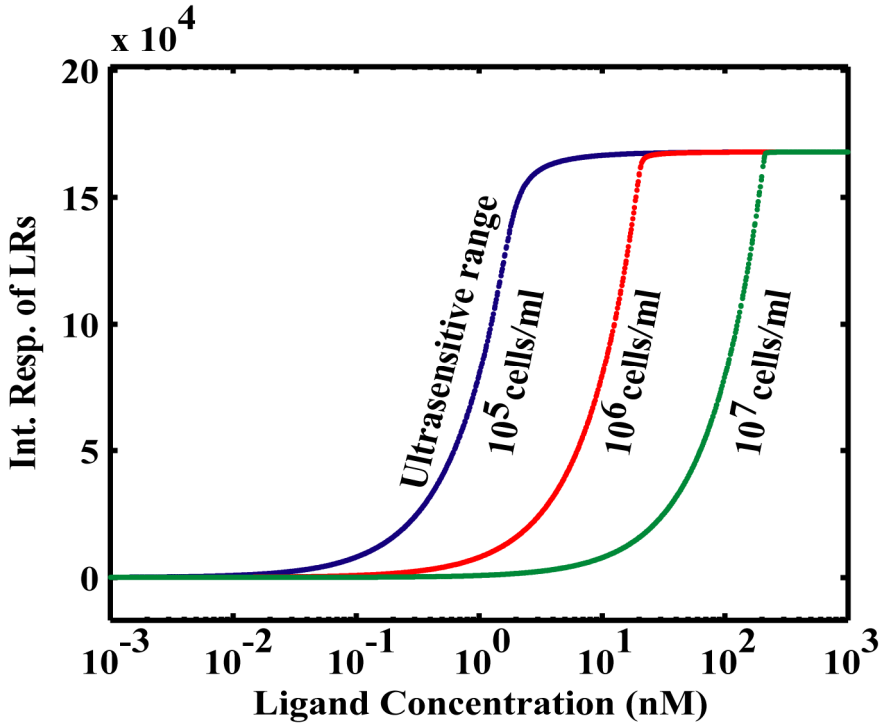

Figure 2.4 Integrated response curves of surface ligand-receptor complex (*LRs*) in different cell density cultures with different ligand concentrations

We also studied the robustness of the dose-response curve on the variation of binding affinity for the ligand-receptor interaction in different cell density cultures. In order to investigate this issue, we simultaneously changed the dose of ligand and the binding affinity. The simulation result indicates that cells have different robustness of dose-response curve upon the variation of binding affinity in different cell density cultures. The higher of the cell density (the smaller of V_{extra} value), the more robust the dose-response curves of *LRs* upon the variation of binding affinity (Figure 2.5A-2.5C).

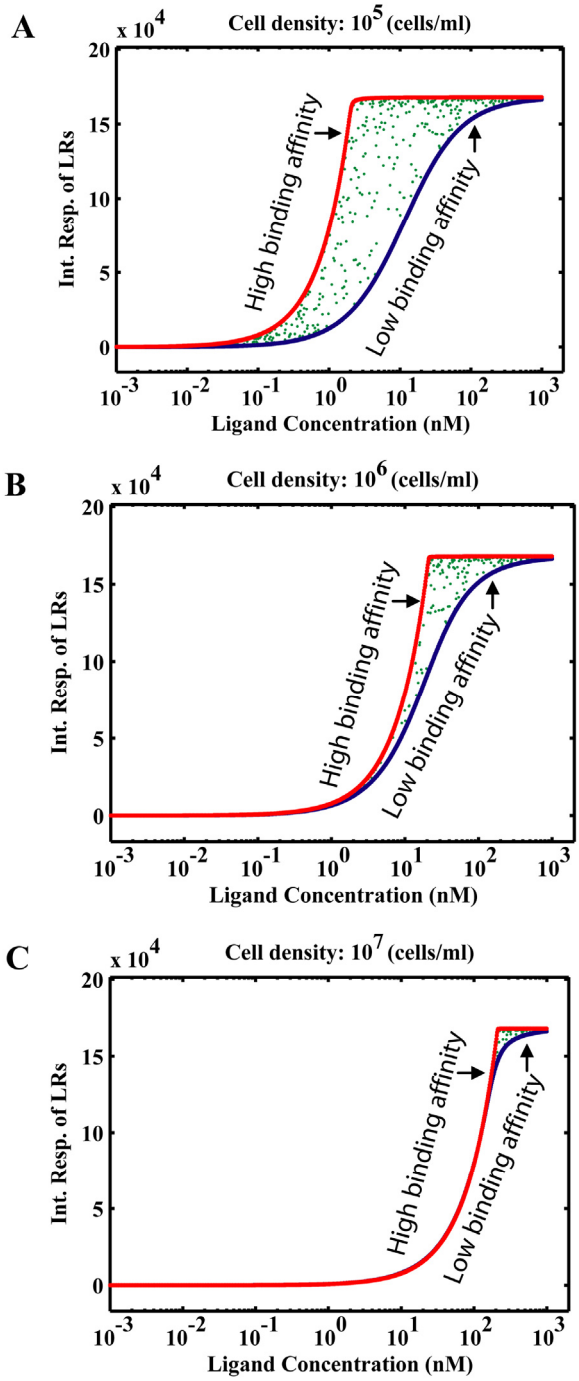

Figure 2.5 Robustness of dose-response curves on the variation of binding affinity in different cell density (*D*) cultures. (A) $D=1\times10^{5}$ cells/ml ($V_{extra}=1\times10^{-8}$ Liter/cell). (B) $D=1\times10^{6}$ cells/ml ($V_{extra}=1\times10^{-9}$ Liter/cell). (C) $D=1\times10^{7}$ cells/ml ($V_{extra}=1\times10^{-10}$ Liter/cell). The red line and blue line represent the boundary dose-response curve of high and low binding affinity, respectively. The green dots represent signal response for the simultaneous variations of ligand concentration and binding affinity.

2.4.2 Cell signal is controlled by the ratio between ligand number and surface receptor numbers per cell

For a certain type of cell, the cell density of the cultures and doses of ligand are usually varied within a wide range under different experimental conditions. We next systematically analyzed which factor determines the signal response when the ligand concentration and cell density are simultaneously varied in the general receptor trafficking network (M1). The result indicates that neither the ligand concentration nor the cell density is the key factor to control the signal response (Figure 2.6A-2.6B).

However, when the correlation of the signal response and the ratio between the ligand number and surface receptor number per cell was investigated, it is surprising that the receptor trafficking network shows a sigmoid response of *LRs* to the ratio between the ligand number and surface receptor number per cell (Figure 2.6C). The ratio between the ligand number and surface receptor number per cell (*RLRs*) is defined as:

$$RLRs = \frac{L_0 \times V_{extra} \times N_{av}}{N_c \times N_r} \tag{2.12}$$

where L_0 is the initial concentration of the ligand in the medium; V_{extra} is the volume of extracellular medium per cell; N_c is the total number of the cells in the culture; N_{av} is the Avogadro's number, 6.02×10^{23}; N_r is the total number of surface receptors per cell. The *RLRs* ratio reflects the ratio between the average available ligand number in medium for each cell and the number of surface receptors per cell.

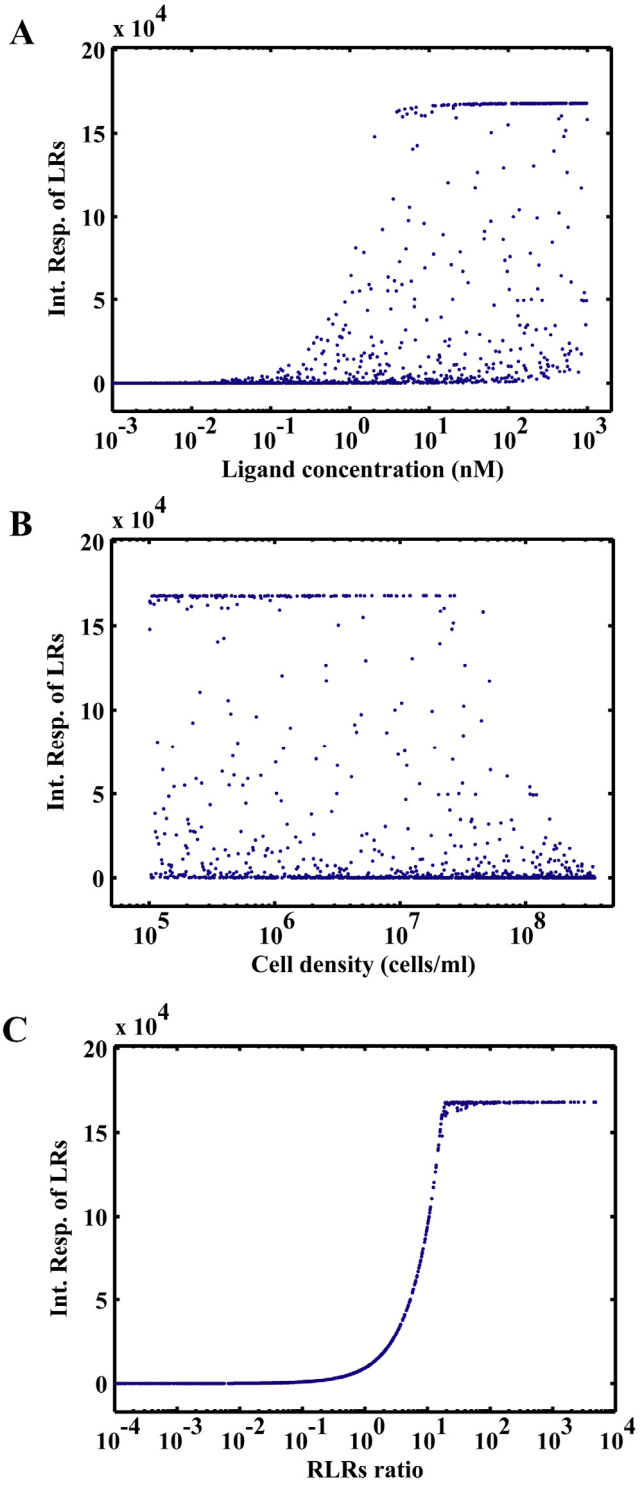

Figure 2.6 Dependence of signal response on ligand concentration, cell density, and *RLRs* ratio. (A) The dependence of integrated signal response of *LRs* on ligand concentration. (B) The dependence of integrated signal response of *LRs* on cell density. (C) The dependence of integrated signal response of *LRs* on *RLRs* ratio.

The next question is whether the sigmoid signal response to the *RLRs* ratio is a general principle in all the receptor trafficking networks (M1-M8 in Figure 2.1). To answer this question, it is necessary to investigate the signal responses of the activated surface ligand-receptor complex (*LRs*) and the internalized ligand-receptor complex (*LRi*), which are coupled to the signal response in different signaling pathway. As shown in Figure 2.7A-2.7B, the sigmoid signal response to the *RLRs* ratio generally occurs in all the investigated receptor trafficking networks. Comparing the relationship between signal response and *RLRs* ratio in the different receptor trafficking networks, the recycling step of empty receptor leads to a higher maximum response of *LRs* and *LRi* (M1 versus M5, M2 versus M6, M3 versus M7 and M4 versus M8 in Figure 2.7A-2.7B). Furthermore, recycling of occupied receptor also increases the maximum response of *LRs*, but it has less effect on the maximum response of *LRi* (M1 versus M2, M3 versus M4, M5 versus M6 and M7 versus M8 in Figure 2.7A-2.7B).

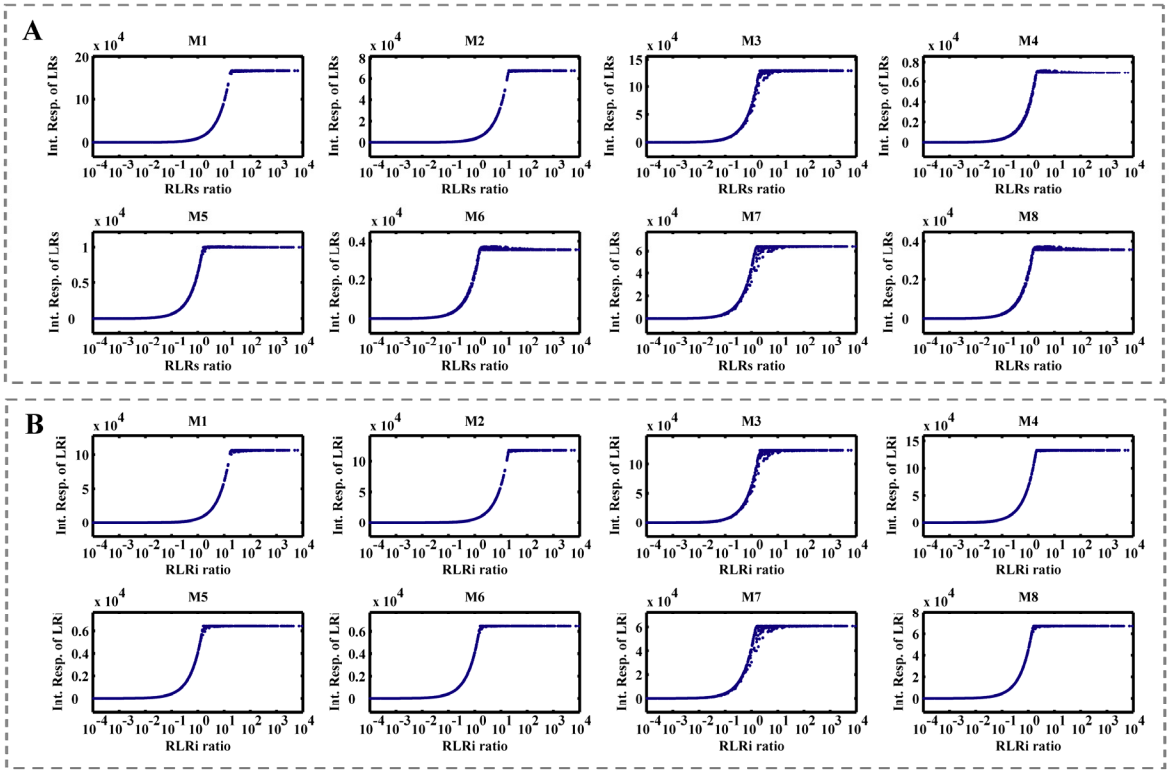

Figure 2.7 Dependence of signal response of different receptor trafficking networks on *RLRs* ratio. (A) Relation of integrated signal response of *LRs* and *RLRs* ratio. (B) Relation of integrated signal response of *LRi* and *RLRs* ratio.

2.4.3 Comparison of model prediction with experimental observations

Finally, the model predictions were compared to the experimental data. Because the EGF signaling pathway has been well studied and documented for more than 20 years, we compared the reported experimental results on the EGF signaling response to different doses of EGF (Clayton et al., 2005; Knauer et al., 1984; Sawano et al., 2002). Although the signal responses in these experiments are measured by different methods and in different systems, all of them indicated that EGF receptor trafficking networks have sigmoid responses to the dose of EGF. On the other hand, the dose-response curve of EGF signal investigated by Knauer et al. also indicated that the same type of cells have different signal responses to the same concentration of EGF in large and small volumes of medium (Knauer et al., 1984) (Figure 2.8B). The dose-response curve of EGF obtained from microfluidic channel was also found to be shifted to the right compared to that from normal culture dish in the work of Sawano et al. (Sawano et al., 2002).

No reasons were proposed to explain that the same type of cells has different responses to the same of dose of EGF. The finding of this study suggests an explanation for the different dose-response curves of EGF for the same type cells in different systems. According to the work by Knauer et al. (Knauer et al., 1984), the difference between the two experiments is the different volume of medium, which means the cell densities are different in the two experiments. The decrease of medium volume corresponds a smaller value of the extracellular medium volume per cell (V_{extra}), which leads to the reduction of the RLRs ratio and then reduces the signal response to EGF. Therefore, the dose-response curve obtained from small medium volume (high cell density) shift to the right compared to that from large medium volume (low cell density) (Figure 2.8A). The same interpretation might be also applied to explain the difference in the work of Sawano et al. (Sawano et al., 2002).

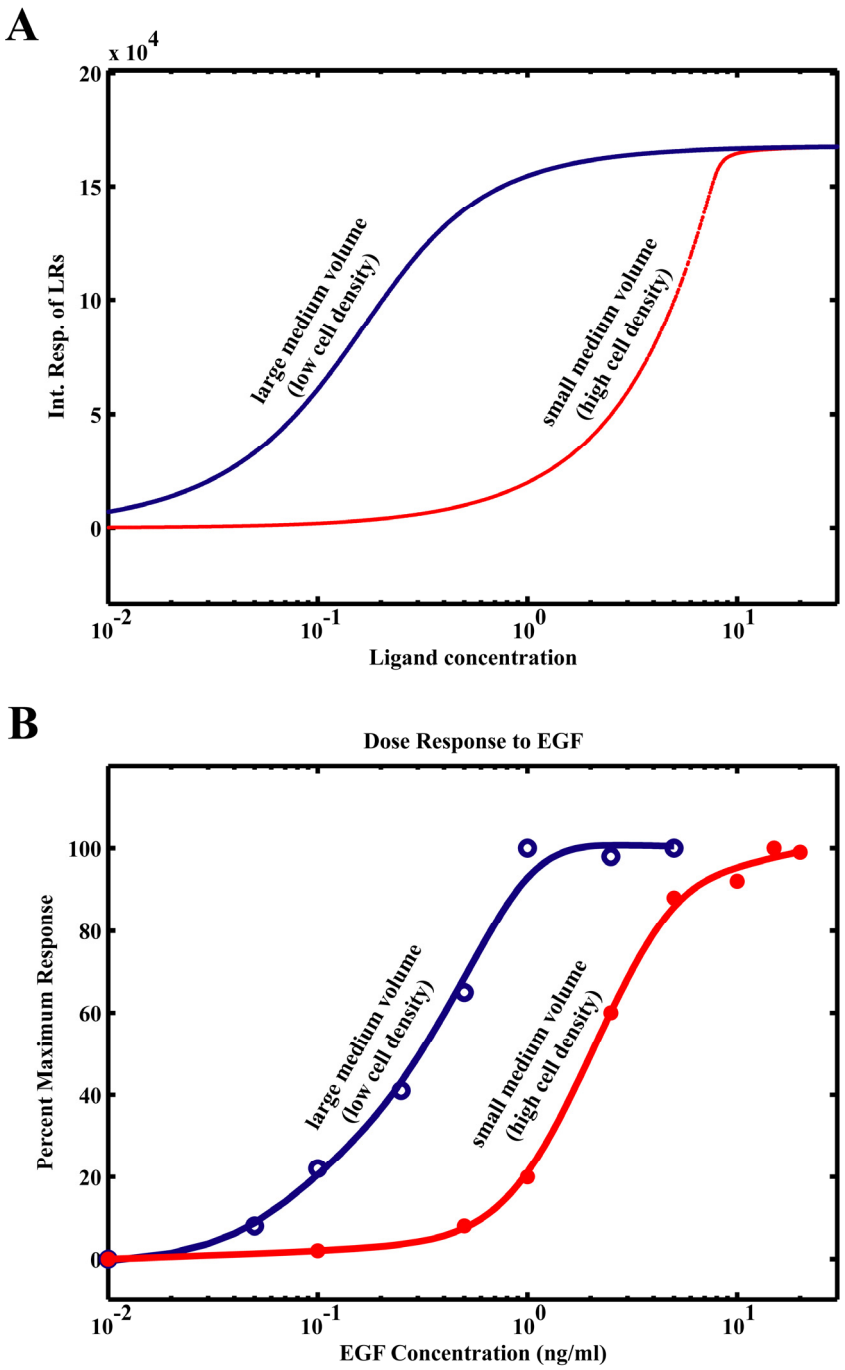

Figure 2.8 Comparison of model predictions to experimental data of EGF dose responses. (A) Model predictions of dose response of signal in small and large medium volumes. (B) Maximum signal responses to EGF in small and large medium volumes. Other conditions of the experiments are the same except the different volume of medium. Panel B of this figure is generated according to the experimental data obtained by Knauer et al (Figure 1B in (Knauer et al., 1984))

2.5 Discussion

Our analysis shows that there is a sigmoid response of activated receptor to the ratio between the ligand number and surface receptor number per cell (*RLRs* ratio) in receptor trafficking networks. The conclusion drawn here doesn't depend on specific arrangements such as specifically assumed receptor trafficking topology, the assumption of the constant or decaying ligand concentration or specific cell density in the culture. Therefore, we propose here that the sigmoid signal response is potentially existent in the receptor trafficking networks. Many factors contribute to the *RLRs* ratio, which enable cells to have diverse mechanisms to control the signal response. One common way of modifying the *RLRs* ratio is the generation of ligand concentration gradients during cell development. The gradient of ligand provides a powerful mechanism for specifying cell fate in developmental biology (Ashe and Briscoe, 2006; Gurdon and Bourillot, 2001). The result in this work suggests that the ligand gradients cover the ultrasensitive range of the sigmoid dose-response curve (Figure 2.4). If the concentrations of the ligand gradient are out of ultrasensitive range, cells can not sense the difference in the ligand concentrations, which might make cells unable to specify different gene expressions and development. Moreover, the signal response of receptor trafficking networks will be saturated when the ligand concentration is larger than a certain threshold. The cell density of the culture for a certain type of cell can change saturation threshold of the dose-response curve (Figure 2.4).

In general, the purpose of this study is not to explain the signal responses in any specific receptor trafficking networks, but rather to investigate the underlying principles for the general responses of receptor trafficking networks. We also want to point out that the mathematical model shown in this study is based on the step signal/ligand scenario, Therefore, the simulations and results shown in this study are relevant to the following cases: the paracrine systems or in experimental scenarios where soluble growth factors are added to the extracelluar medium

In summary, cellular signal response is potentially controlled by ratio between ligand number and surface receptor number per cell. Sigmoid signal response is a general principle in different receptor trafficking networks. It is expected that it will not be possible to accurately predict cell signal response in receptor trafficking networks if we take for granted that the ligand concentration in the medium is constant or follow a decaying function over time. Therefore, to fully understand the complexity of cellular signaling pathways it will be

necessary to consider the ligand concentration and cell density which contribute to the variation of the ratios between ligands and surface receptors per cell.

Chapter 3 SBML-PET: a Systems Biology Markup Language (SBML) based Parameter Estimation Tool

3.1 Introduction

The aim of parameter estimation for biochemical network models is to find the most feasible parameters that reproduce experimental results according to a given set of experimental data. Several efforts have been dedicated to solve this problem. Some researchers focus on the development of optimization methods (Gadkar et al., 2005; Klipp and Heinrich, 1994; Kremling et al., 2004; Matsubara et al., 2006; Mendes and Kell, 1998; Runarsson and Yao, 2000; Tsai and Wang, 2005; Zak et al., 2003), while others pay attention to making the methods available for the users. For example, Grid Cellware (Dhar et al., 2005) and OBIYagns (Kimura et al., 2004) are grid-based tools with parameter estimator. Gepasi (Mendes, 1993) and Copasi (Hoops et al., 2006), both complex pathway simulators, enable parameter estimation by using data from time course or steady state experiments. The C library libSRES (Ji and Xu, 2006) is composed of a parameter estimator mainly implementing stochastic ranking evolution strategy (SRES) algorithm.

Although there are some software packages with parameter estimation facility, it is still hard for systems biologists to apply them to their specific problems. First, none of the currently existing software package supports models in which events have been defined, although events are very common and important for biological experiments. Most experimental data refer to a specific event in the experiment like changing signal source or strength at a specific time, or blocking protein synthesis during the experiment. Currently available software packages are difficult to deal with such cases. Secondly, sometimes it is impossible to directly measure the concentration of the components in the model. Instead, only the sum amount of several components or the relative change of a component can be measured. Parameter estimation software should support different kind of experimental data, for example, measurement error in the experimental data, normalized data or some mathematical expression with the concentration of the components.

In order to satisfy the specific customized requirements for parameter estimation, we developed an SBML based Parameter Estimation Tool (SBML-PET). SBML-PET is designed to perform parameter estimation for biological models including signaling pathways, gene regulation networks and metabolic pathways. SBML-PET has a feature of supporting the models including events. It also supports a variety of data from different experimental conditions and the data can be expressed with common mathematical expression (for example, trigonometric and transcendental functions).

3.2 Overview of SBML-PET

SBML-PET is a Systems Biology Markup Language (SBML) based Parameter Estimation Tool. SBML-PET supports import and export of models in the SBML format. Currently, it can run in Linux and with Cygwin in Windows. SBML-PET runs in a command interactive mode environment. It is easy to install and to use. Following the guide information printed on the screen, the user can easily finish the parameter estimation work for the SBML model. Before implementing SBML-PET, the user needs to prepare the data file following the instructions for the data file. Detailed information about the preparation of data file will be described later.

3.2.1 Features of SBML-PET

The features of SBML-PET are summarized in the following:

- SBML-PET supports model import and export in SBML format, a widely accepted standard for the exchange of biochemical network models. All estimated parameters will be saved in a new SBML file, which can be imported to other SBML supported software packages.
- SBML-PET supports events that describe the discontinuous state changes in the model.
- Frequently, experimental data is produced under different conditions. SBML-PET can estimate the parameters using different types of data.
- Concentration measurements are often indirect. SBML-PET supports a variety of experimental data, for example, normalized data or common mathematical expression of the molecule concentrations.

- SBML-PET supports common mathematical expressions for the qualitative and quantitative description of the model, for example, the constraints for parameters or their combinations.

- SBML-PET also supports the standard deviation for the data and the noise or measurement error existed in the experiments

3.2.2 System and package requirements for SBML-PET

In order to install and use SBML-PET, the user needs to install the following packages according to the different platforms:

- Linux: GCC library. Although SBML-PET has been tested on gcc 2.9.5, gcc-3.3.1, gcc-3.3.5, gcc-4.03, Gcc-3.3.1 or higher version of GCC is recommended.

- Windows: Install Cygwin and select all the packages (gcc-core, gcc-g++, gcc-g77 etc.) for gcc in "Devel Category". Cygwin install information can get from http://www.cygwin.com/.

3.2.3 Installation of SBML-PET in Linux

- Step 1: Download the SBML-PET ZIP file from the website of Kinetic Modeling Group in Max Planck Institute for Molecular Genetics (http://sysbio.molgen.mpg.de/SBML-PET/Download.html).

- Step 2: Unzip the zip file and it will automatically create a folder "SBML-PET" with several subfolders and files inside.

- Step 3: Setup the LD_LIBRARY_PATH variable. You can do this by adding the following line at the end of /etc/profile (for all users) or $HOME/.bashrc file (for current user). $SBML-PET_PATH is the full path for the directory including the unzipped SBML-PET files.

  export LD_LIBRARY_PATH=$SBML-PET_PATH/lib:$LD_LIBRARY_PATH.

- Step 4: start a new shell and change to the directory including the unzipped SBML-PET files. Execute SBML-PET by typing ./SBML-PET

3.2.4 Installation of SBML-PET in Cygwin of Windows

- Step 1: Download the SBML-PET ZIP file from the website of Kinetic Modeling Group in Max Planck Institute for Molecular Genetics (http://sysbio.molgen.mpg.de/SBML-PET/Download.html).

- Step2. Unzip the file into your hard drive that Cygwin can access. It will automatically create a folder "SBML-PET" with several subfolders and files inside.

- Step3. Stat Cygwin and change to the directory including the unzipped SBML-PET files. Execute SBML-PET.exe by typing ./SBML-PET.

3.3 Used Libraries

SBML-PET incorporates Stochastic Ranking Evolution Strategy (SRES) for parameter estimation. SRES is a (μ,λ)-ES evolutionary optimization algorithm that uses stochastic ranking as the constraint handling technique (Runarsson and Yao, 2000). More detailed description of SRES algorithm is provided in Appendix B. Moles et al. compared several common global optimization methods and showed that SRES achieves the best result in the case of a three-step pathway (Moles et al., 2003). The performance of SRES has been extensively tested on many parameter estimation problems and it has good performances for the tested cases. On the other hand, all the solutions of ODE systems in SBML-PET are computed by ODEPACK (Hindmarsh, 1983) which is a collection of FORTRAN ODE solvers.

3.4 Usage of SBML-PET

SBML-PET runs in a command interactive mode environment. Following the guide information shown the screen, the user can easily estimate the parameters for the SBML model.

3.4.1 Start of SBML-PET

The user can run SBML-PET by typing the command "./SBML-PET" in the shell. The welcome information will appear (Figure 3.1).

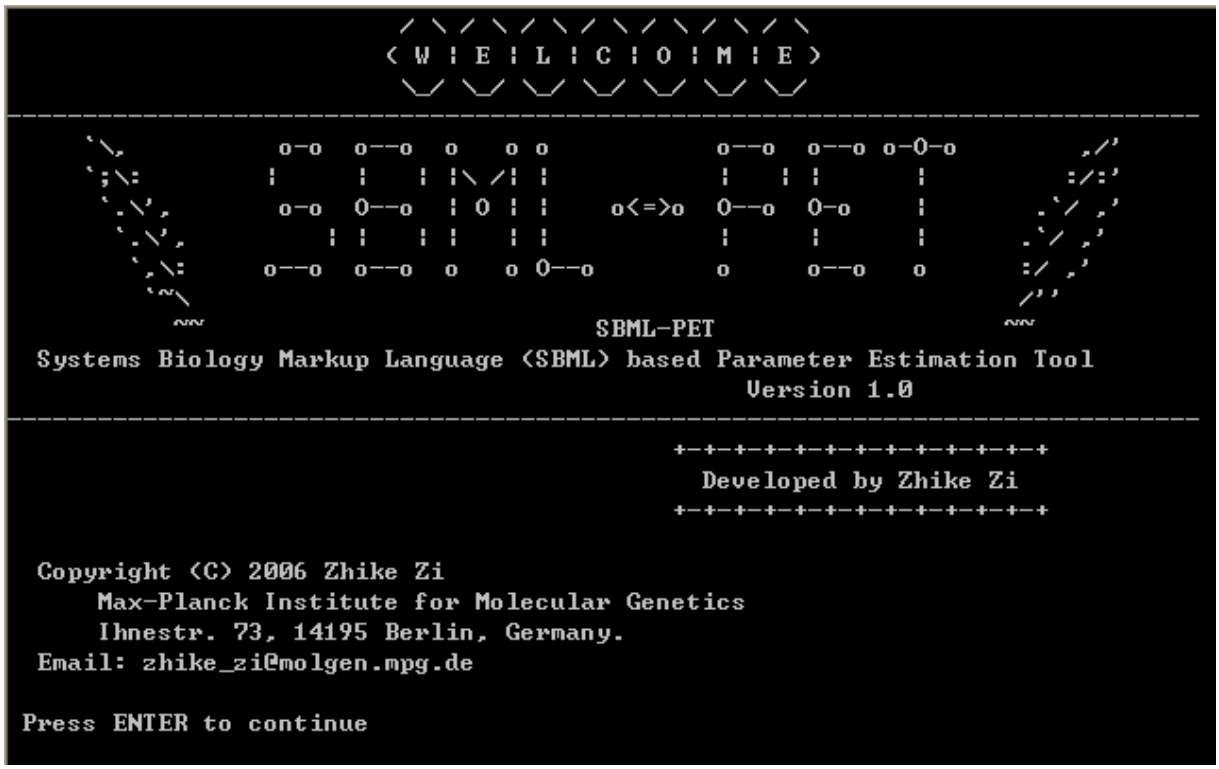

Figure 3.1 Start of SBML-PET

3.4.2 Task and advanced options

The next step is to set the task and advanced options (Figure 3.2). SBML-PET can run parameter estimation and also simulation for the SBML models.

As far as advanced options are concerned, choose “1 Using Default Settings” by typing “1” and press ENTER key to continue if the user is not familiar with the ODE solver and evolutionary algorithm. If the user wants to set the parameters for ODE solver and evolutionary algorithm, type “2” and press ENTER key. The next step is to set some parameters for ODE Solver and evolutionary algorithm. The parameters for advanced options will be explained later in details.

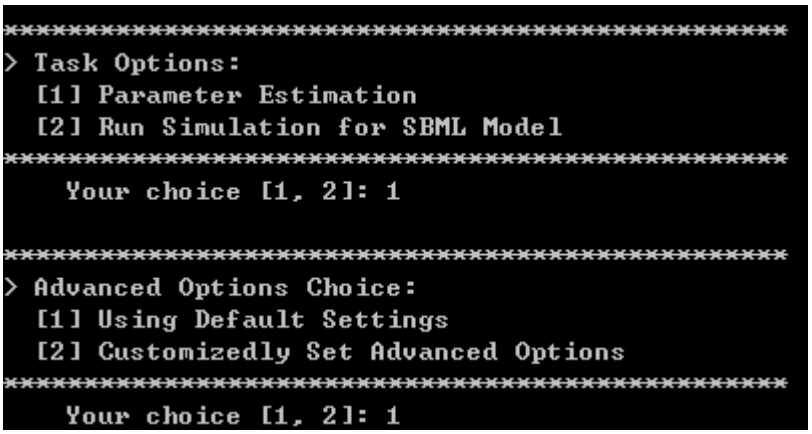

Figure 3.2 Advanced options choice

3.4.3 Import of SBML file

The user can import the model in SBML format. The path of the SBML file should be included (Figure 3.3). Both the absolute path and relative path are acceptable in SBML-PET. SBML-PET will check the validity of the SBML file and display the basic information about the SBML model.

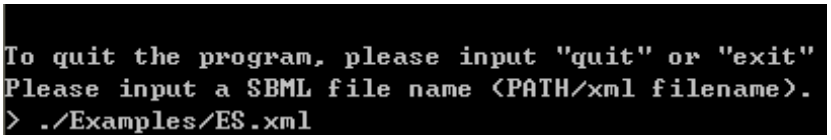

Figure 3.3 Import SBML file

3.4.4 Prepare the data file for parameter estimation

The user can specify the parameters to be estimated and the experimental data in the data file. The following part will explain the data file with more details.

(1) Annotations for Data File

The user can add annotation by using "#" at the beginning of each line (Figure 3.4).

```
#************************************************************
#  Input data file for SBML-PET
#
#  You can add annotation by using "#" at the beginning
#  of the annotation line.
#
#************************************************************
```

Figure 3.4 Annotations in SBML-PET

(2) Information about Parameters to Be Estimated

Copy the information about parameters to be estimated to space after PART I in the data file. This part should include “PARAMETER ID”, “MINIMUM VALUE” and “MAXIMUM VALUE” information (Figure 3.5).

```
#************************************************************
#  PART I: Parameter to be Estimated
#
#  Copy the information for parameters to be estimated
#  to the below
#
#  Attention:
#    Don't insert blank line in the middle of
#    the data for parameters to be estimated.
#************************************************************
  PARAMETER ID      MINIMUM VALUE     MAXIMUM VALUE
           E              0.1             1000
           S              0.1             1000
          k1            0.001              1e3
          k2            0.001              1e3
          k3            0.001              1e3
```

Figure 3.5 Information about parameters to be estimated

The species ID shown in “Parameter ID” corresponds to the initial concentration/amount of the species.

(3) Information about Experimental Data

The information about experimental data is supposed to be in PART II (Figure 3.6).

```
#************************************************************
#  PART II: Experimental Data
#
#  Copy the information for experimental data to the below
#
#  Attention:
#    Don't insert blank line in the middle of
#    the time course data
#************************************************************
TOTAL NUMBER of EXPERIMENTAL CONDITIONS = 1

# data information at Experimental Condition 1
AT CONDITION 1, the TOTAL NUMBER of TIME COURSES is 1
DATA OF TIME COURSE 1 AT CONDITION 1
NUM_SAMPLING_POINTS 10 NUM_EXP_DATA 2
Time            S              SD           P          SD
1            0.924055       0.04620275    4.94519    0.2472595
2            0.268851       0.01344255    7.84506    0.392253
3            0.0963395      0.004816975   9.11061    0.4555305
4            0.0372971      0.001864855   9.63658    0.481829
5            0.0148637      0.000743185   9.85204    0.492602
6            0.00599125     0.000299563   9.93985    0.4969925
7            0.00242596     0.000121298   9.97556    0.498778
8            0.00098412     0.000049206   9.99007    0.4995035
9            0.000399518    1.99759E-05   9.99597    0.4997985
10           0.00016224     0.000008112   9.99836    0.499918
Normalized      0                            0
```

Figure 3.6 Information about experimental data

- Modify the total number of experimental conditions in the following line:

  TOTAL NUMBER of EXPERIMENTAL CONDITIONS = 1

- Modify the time course number and condition number for each time course data:

  AT CONDITION 1, The TOTAL NUMBER of TIME COURSES is 1

- Modify the number of sampling points and number of experimental data for each time course data:

  NUM_SAMPLING_POINTS 10 NUM_EXP_DATA 2

- Input the mathematical expression for each data in line:

  Time S SD P SD

SD is for standard deviation σ or noisy of data. If you have two experimental data for x2 + 2*x3 and k_scale*(x1 + x2 + 2*x3) respectively, you can input it as:

Time x2 + 2*x3 SD k_scale*(x1 + x2 + 2*x3) SD

- Insert the time course data after the mathematical expression of data.

1	0.924055	0.04620275	4.94519	0.2472595
…	…	…		
9	0.000399518	1.99759E-05	9.99597	0.4997985
10	0.00016224	0.000008112	9.99836	0.499918

The data can be directly copied from Excel. If the standard deviation for the data is unknown, set all of them to be 0. The standard deviation information must be provided. Otherwise SBML-PET can not run normally.

- Insert normalized information for the data. The number 0 indicates not normalized data and 1 means normalized data.

Normalized	0	1	0	….

For SD column, there is no normalized information. Therefore, the number of columns in the line of normalized information is different from that of time course data.

(4) Information about Constraints

The information about constraints is provided in the last part of the data file (Figure 3.7). Please pay attention to using the Id (NOT the name) of parameters in the mathematical expression of constraints.

```
#***********************************************************
#  PART III: Constraints Information
#
#  Copy the information for Constraints to the below
#***********************************************************

# Constraints Information at Experimental Condition 1
AT CONDITION 1, the TOTAL NUMBER of CONSTRAINTS is 2
MATH OF CONSTRAINT 1 AT CONDITION 1, k1 - 10*k2
MATH OF CONSTRAINT 2 AT CONDITION 1, k2 - 10*k1
```

Figure 3.7 Information about constraints

The math expression *g(X, p)* corresponds to the constraints with the inequality of

$$g(X, p) \leq 0 \tag{3.1}$$

Where X is the involved initial conditions of the molecule amount/concentration and p is the parameter set. For the example “k1-10*k2” corresponds to “k1-10*k2≤0” and that is k1≤10*k2.

Please DO NOT MODIFY THE RESERVED WORDS AND PHRASES (UPPERCASE) IN THE DATA FILE. The reserved words and phrases are list in the following, which are reserved for SBML-PET:

PARAMETER ID	MINIMUM VALUE
MAXIMUM VALUE	TOTAL NUMBER
EXPERIMENTAL CONDITIONS	AT CONDITION
TIME COURSE	TIME COURSES
NUM_SAMPLING_POINTS	NUM_EXP_DATA
MATH	DATA
CONSTRAINT	CONSTRAINTS

3.4.5 Input the data file for SBML-PET

If the data file is prepared, input the name of the data file and press ENTER key to continue (Figure 3.8).

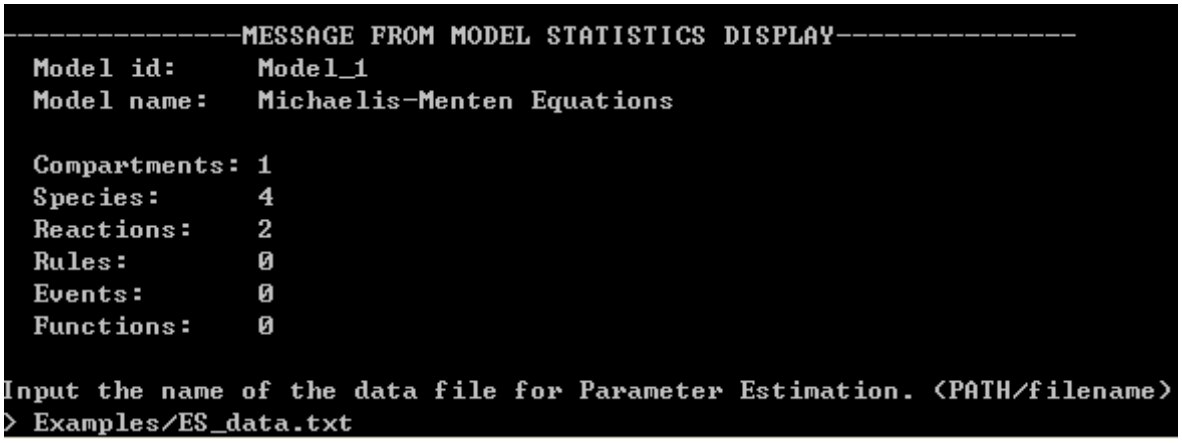

Figure 3.8 Input data file

3.4.6 View result

The user can view the parameter estimation result including used time and best estimated parameters for each evolution generation while SBML-PET is running (Figure 3.9).

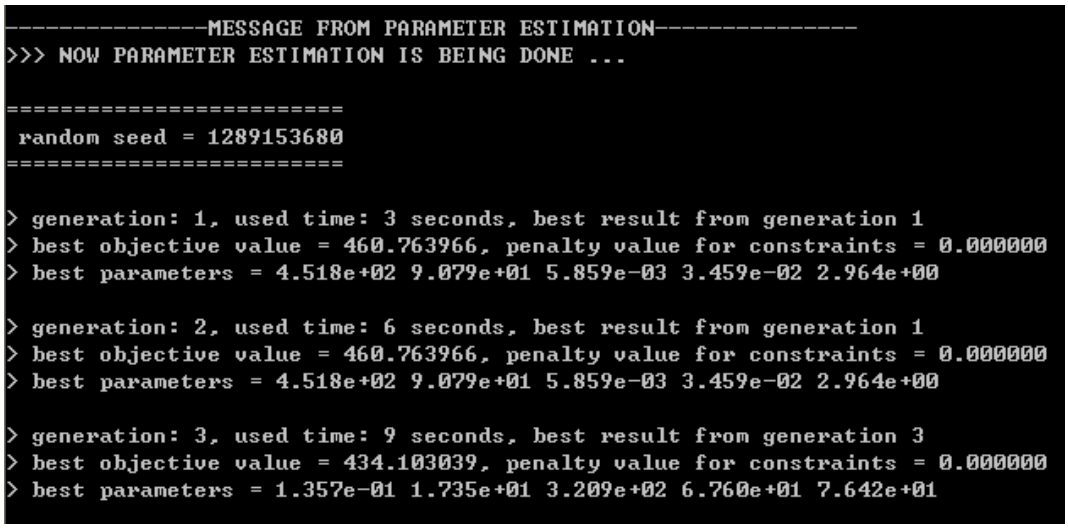

Figure 3.9 Result display in SBML-PET

The estimated parameter values are saved in a SBML file called "BestFitnessSBML.xml", which can be imported to other simulation software packages such as Copasi, CellDesigner and so on. If the result is satisfied, the user can press "Ctrl + C" to terminate the program. The result will be automatically stored in the directory of result. Otherwise, the program will stop when the last generation is finished.

3.4.7 Explanation for multi experimental conditions

If the experimental data are obtained from different experimental conditions, the user need modify some parameters defined in the model to distinguish the difference of the conditions.

After the SBML file of the general model for different conditions is input, a file named "ExpConditionsData.txt" will be produced at the "./temp" directory of SBML-PET. For different experimental conditions, the possible varied parameters values will be listed in this file. The user can modify the corresponding data for different conditions and save the file, then press ENTER key to continue (Figure 3.10).

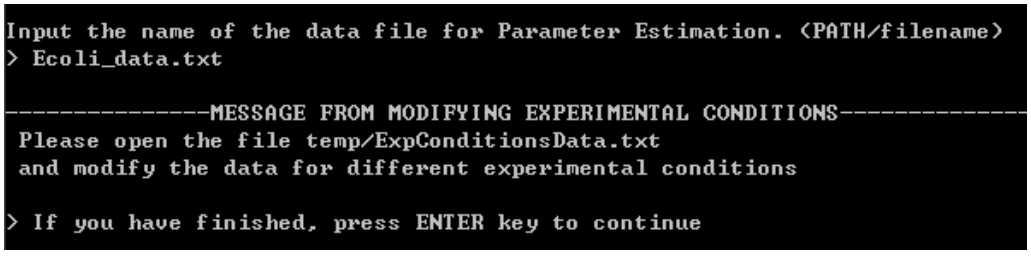

Figure 3.10 Modify experimental condition data

The value of the species listed in the file of "ExpConditionData.txt" indicates the corresponding initial concentration/amount of the species. If the user wants to use some parameters to distinguish different experimental conditions, please define these parameters as global parameters rather than local parameters in the model.

3.4.8 Advanced options for ODE solver

Currently, SBML-PET uses two ODE solvers from ODEPACK (Hindmarsh, 1983) to solve the ODE systems:

- LSODA solves ODE systems with automatic selection between non stiff and stiff methods. It uses the non stiff method initially, and dynamically monitors data in order to decide which method to use.

- LSODAR is a variant of LSODA with a root finding capability added. Thus it solves problems with automatic method selection, and at the same time, it finds the roots of a set of given functions. This is useful for event location.

If there is no event in the SBML model, LSODA ODE solver is used. Otherwise, LSODAR ODE solver is chosen. For the ODE Solvers, the user can set the following parameters, which is shown in Figure 3.11.

```
*****************************************************
> Setting ODEPACK Parameters (Default values):
    Relative Tolerance Parameter (scalar): 1e-06
    Relative Absolute Parameter (scalar): 1e-06
    Maximum Number of Steps allowed during one call: 10000
*****************************************************
Please input Relative Tolerance Parameter (scalar)
> 1e-6
Please input Absolute Tolerance Parameter (scalar)
> 1e-12
Please input Maximum Number of Steps allowed during one call to the solver.
> 5000
```

Figure 3.11 Parameters setting for ODE solver

- Relative Tolerance Parameter (RTOL). RTOL is used to control the estimated local error for ODE solver. The default value of RTOL in SBML-PET is 1×10^{-6}.

- Absolute Tolerance Parameter (ATOL). The estimated local error in Y(i) will be controlled so as to be roughly less (in magnitude) than

  EWT(i) = RTOL×ABS(Y(i)) + ATOL

  The default value of ATOL in SBML-PET is 1×10^{-6}.

- Maximum Number of Steps (MXSTEP). It is the maximum number of (internally defined) steps allowed during one call to the solver. The default value is 10000.

3.4.9 Advanced options for evolutionary algorithm

SBML-PET uses Stochastic Ranking Evolution Strategy (SRES) to run parameter estimation (Runarsson and Yao, 2000). It provides the following parameter setting (Figure 3.12).

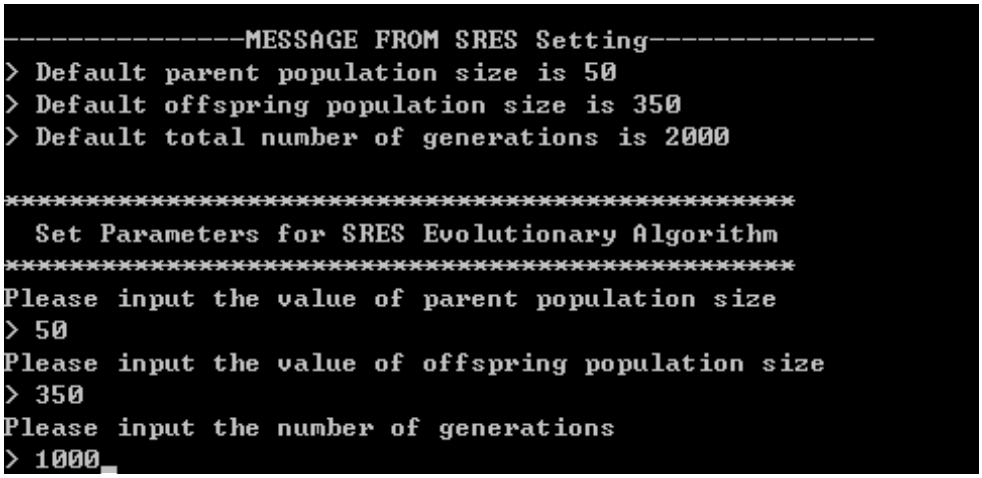

Figure 3.12 Parameters setting for SRES evolutionary algorithm

- Parent population size. The default value is 50. Here, parent population size is recommended to be in the range of 20~100.

- Offspring population size. The default value is 350.

- Total number of generations. This is the user-defined number of evolution generations. The default value is 2000.

The ratio between parents and offspring is about 1/3~1/10 depending on specific cases. According to the previous experiences reported, the ratio around 1/7 is good for most of the

problems. However, it is not 100% correct for every problem. The user needs to do some preliminary experiments to achieve the best result in practice.

3.5 Application of SBML-PET

3.5.1 Example of Michaelis-Menten equations

The well-known Michaelis-Menten equations describe enzyme kinetics, which include four molecular species, namely the enzyme, *E*, the substrate, *S*, the product, *P* and the intermediate, *ES*. The reactions are list below:

$$E + S \underset{k_2}{\overset{k_1}{\Leftrightarrow}} ES \xrightarrow{k_3} E + P \tag{3.2}$$

The ODE system consists of 4 ODE.

$$\frac{d[E]}{dt} = k_2[ES] + k_3[ES] - k_1[E][S] \tag{3.3}$$

$$\frac{d[S]}{dt} = k_2[ES] - k_1[E][S] \tag{3.4}$$

$$\frac{d[ES]}{dt} = k_1[E][S] - k_2[ES] - k_3[ES] \tag{3.5}$$

$$\frac{d[P]}{dt} = k_3[ES] \tag{3.6}$$

Table 3.1 Range of parameters to be estimated

Parameter ID	Minimum Value	Maximum Value
E	0.1	1000
S	0.1	1000
k1	0.001	1000
k2	0.001	1000
k3	0.001	1000

In order to clearly demonstrate the ability of SBML-PET, we assume that the true values of $k_1 = 1$, $k_2 = 1$, $k_3 = 1$ and the initial concentration of $[E] = 10$, $[S] = 10$, $[ES] = 0$, $[P] = 0$. Then we run SBML-PET to estimate the values of k_1, k_2, k_3 and the initial concentration of E and S within the following range (Table 3.1).

There are 5 parameters to be estimated. The initial concentration of E and S cover four magnitudes, while the value of kinetic parameters k_1, k_2 and k_3 cover six magnitudes.

The following pseudo experimental data is used for parameter estimation (Table 3.2). Standard variation for the data is set to be 5% of the each data.

Table 3.2 Pseudo experimental data for Michaelis-Menten equations

Time	S	SD	P	SD
1	0.924055	0.04620275	4.94519	0.2472595
2	0.268851	0.01344255	7.84506	0.392253
3	0.0963395	0.004816975	9.11061	0.4555305
4	0.0372971	0.001864855	9.63658	0.481829
5	0.0148637	0.000743185	9.85204	0.492602
6	0.00599125	0.000299563	9.93985	0.4969925
7	0.00242596	0.000121298	9.97556	0.498778
8	0.00098412	0.000049206	9.99007	0.4995035
9	0.000399518	1.99759E-05	9.99597	0.4997985
10	0.00016224	0.000008112	9.99836	0.499918
Normalized	0		0	

The number of constraints for SBML is set to be 1 and the mathematical expression is:

k1 – 10 *k2 which means that k1<10*k2.

According to the description in section "3.4.4 Prepare the Data File for Parameter Estimation", the data file for parameter estimation is shown in Figure 3.13.

```
**********************************************************
#  PART I: Parameter to be Estimated
#**********************************************************
 PARAMETER ID     MINIMUM VALUE    MAXIMUM VALUE
      E                 0.1              1000
      S                 0.1              1000
      k1                0.001            1000
      k2                0.001            1000
      k3                0.001            1000

#**********************************************************
#  PART II: Experimental Data
#**********************************************************
TOTAL NUMBER of EXPERIMENTAL CONDITIONS = 1
# data information at Experimental Condition 1
AT CONDITION 1, the TOTAL NUMBER of TIME COURSES is 1
DATA OF TIME COURSE 1 AT CONDITION 1
NUM_SAMPLING_POINTS 10 NUM_EXP_DATA 2
Time        S              SD             P          SD
1        0.924055       0.04620275      4.94519    0.2472595
2        0.268851       0.01344255      7.84506    0.392253
3        0.0963395      0.004816975     9.11061    0.4555305
4        0.0372971      0.001864855     9.63658    0.481829
5        0.0148637      0.000743185     9.85204    0.492602
6        0.00599125     0.000299563     9.93985    0.4969925
7        0.00242596     0.000121298     9.97556    0.498778
8        0.00098412     0.000049206     9.99007    0.4995035
9        0.000399518    1.99759E-05     9.99597    0.4997985
10       0.00016224     0.000008112     9.99836    0.499918
Normalized      0              0

#**********************************************************
#  PART III: Constraints Information
#**********************************************************
# Constraints Information at Experimental Condition 1
AT CONDITION 1, the TOTAL NUMBER of CONSTRAINTS is 1
MATH OF CONSTRAINT 1 AT CONDITION 1, k1 - 10*k2
```

Figure 3.13 Parameter estimation data file for Michaelis-Menten equations model in SBML-PET

The progress of the values of objective function (cost function) is shown in Figure 3.14. The best fit result is obtained within 5 minutes by SBML-PET.

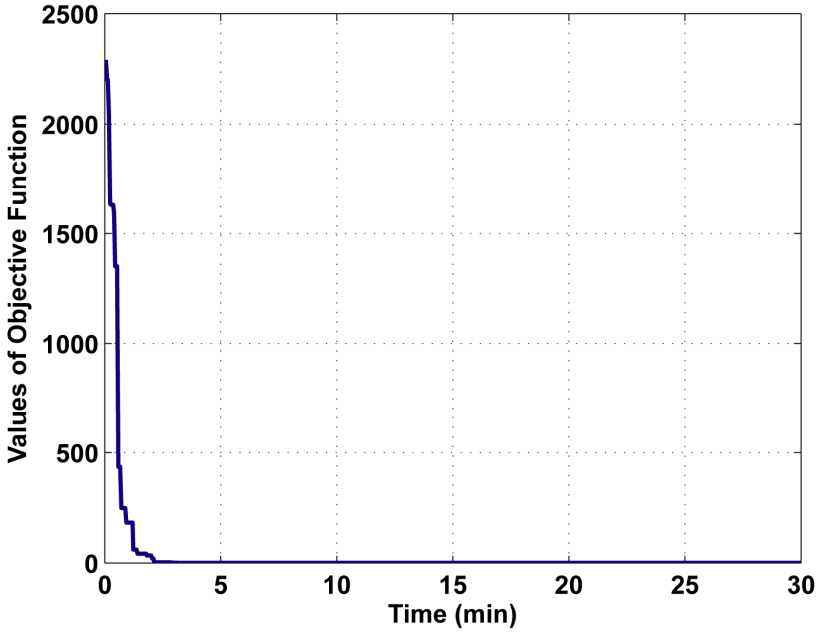

Figure 3.14 Progress of the objective function values

3.5.2 Implementation of SBML-PET

The implementation of SBML-PET for the example of Michaelis-Menten equations is listed in the following.

- Step 1: Change to the directory including SBML-PET files and type "./SBML-PET" command in the shell. After reading the welcome information, press ENTER key to continue.

- Step 2: For the task options, input 1 to do parameter estimation. For the advanced options, input 1 (using default settings) and press ENTER key to continue.

- Step 3: Import SBML file by inputting ./Examples/ES.xml and press ENTER key to continue. SBML-PET will check the validity of the SBML file and display the model statistic information.

- Step 4: Input the path and name of the data file for parameter estimation by type ./Examples/ES_data.txt in the shell and press ENTER key to continue.

The user can monitor the real time result in SBML-PET. When the last generation is finished, the best result for the estimated parameter values will be written in ./result/ BestFitnessSBML.xml SBML file.

3.5.3 JAK-STAT model (real experimental data and events included)

In the following, a test of the tool property is present for an accepted model from the literature. Swameye *et al.* developed a simple JAK-STAT model by database modeling (Swameye et al., 2003). The parameters are estimated according to real experimental data with noisy measurement. Here we make a small modification to this model and estimated the relative parameters with SBML-PET. The JAK-STAT pathway in this model includes 4 proposed reactions. JAK protein binds to the receptor called Erythropoietin receptor (EpoR) and cause the phosphorylation of STAT5 protein. Phosphorylated STAT5 protein then forms a dimmer and import to the nucleus. In the nucleus, phosphorylated STAT5 dimmer will be dephosphorylated and forms a STAT5 monomer and exports back to cytoplasm (Figure 3.15).

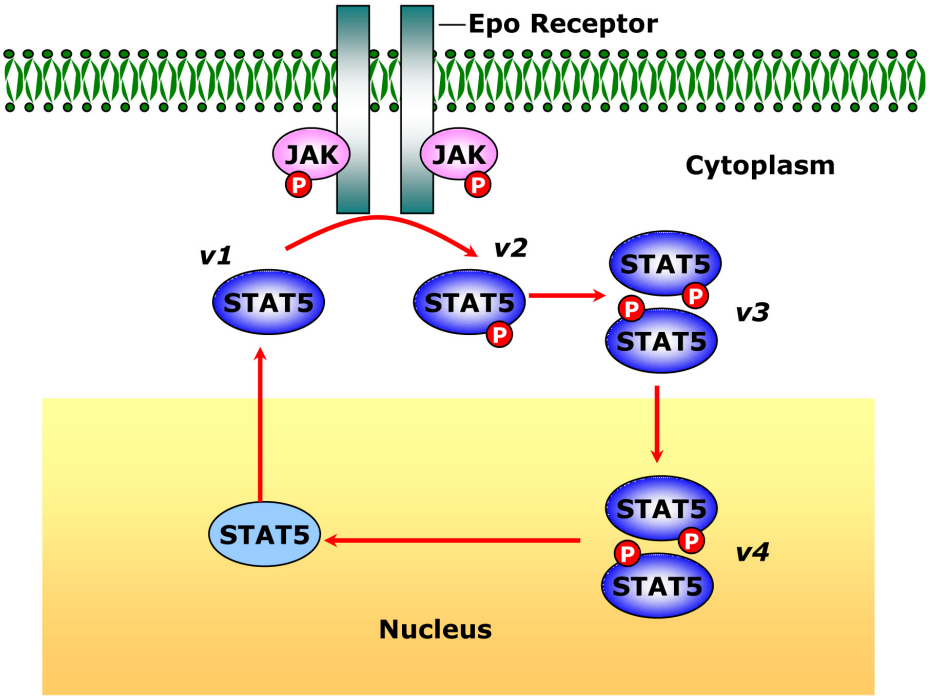

Figure 3.15 Simplified JAK-STAT pathway

The ODE system of modified simple JAK-STAT model is listed below:

$$\frac{dv1}{dt} = -r1 * v1 * D + 2 * r4 * v4 \tag{3.7}$$

$$\frac{dv2}{dt} = +r1 * v1 * D - 2 * v2 * v2 \tag{3.8}$$

$$\frac{dv3}{dt} = -r3 * v3 + v2 * v2 \tag{3.9}$$

$$\frac{dv4}{dt} = +r3 * v3 - r4 * v4 \tag{3.10}$$

$$vi = k2 * xi \quad i \in \{1, 2, 3, 4\} \tag{3.11}$$

$$r1 = k1 / k7,\ r3 = k3 \text{ and } r4 = k4 \tag{3.12}$$

The initial conditions for the model are

$$v1 \text{ (to be estimated)} \quad v2 = 0; \quad v3 = 0; \quad v4 = 0 \tag{3.13}$$

The available experimental data are the total amount of phosphorylated STAT5 and total STAT5 (phosphorylated and unphosphorylated forms) in the cytoplasm, which can be expressed as:

$$y1 = v2 + 2 * v3 \tag{3.14}$$

$$y2 = v1 + v2 + 2 * v3 \tag{3.15}$$

Similar as the original model, we assume that there is a delay time (*tau*) for the phosphorylated STAT5 dimmer to return the cytoplasm because it should be dephosporylated first and then export to the cytoplasm. We incorporate this information as an event in the SBML model, which is:

$$\begin{aligned} &\text{Event trigger : } time < tau \\ &\text{Event Assignment : } r4 = 0 \end{aligned} \tag{3.16}$$

The variables *vi* (*i* = 1, 2, 3, 4) are the scaled value of the concentration of the species STAT5, phosphorylated STAT5 monomer, phosphorylated STAT5 dimer in the cytoplasm，phosphorylated STAT5 dimmer and STAT5 monomer in the nucleus respectively. *D* is the

input function of the receptor activity for the ODE system, which is approximated from the experimental data (Figure 3.16A).

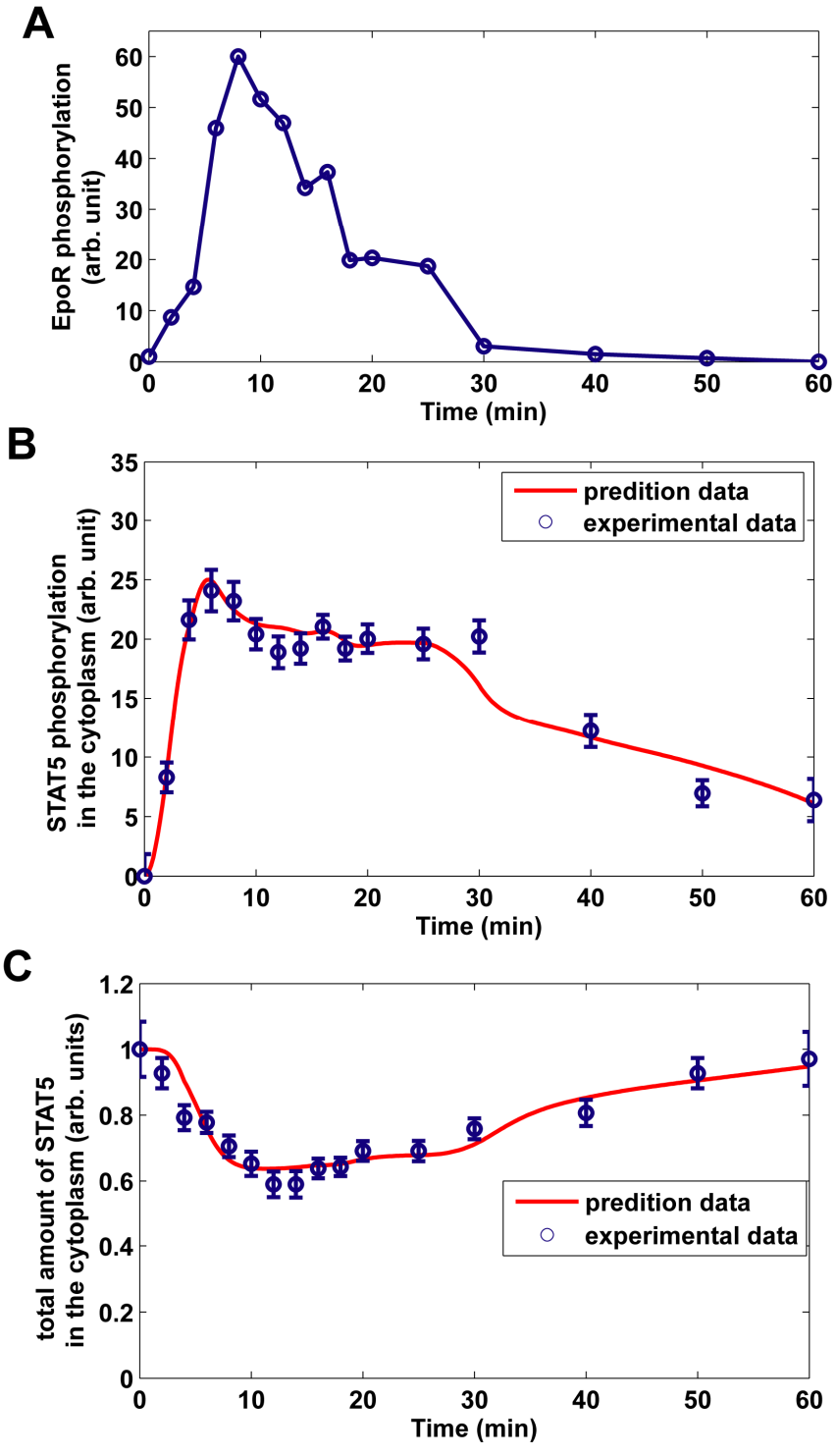

Figure 3.16 Parameter estimation result for JAK-STAT model. (A) The receptor phosphorylation profile (input function of the model *D*). (B) Goodness of model fit to total phosphorylated STAT5 in cytoplasm. (C) Goodness of model fit to total STAT5 in cytoplasm.

We used SBML-PET to estimate the parameters *v*1, *r*1, *r*3, *r*4 and *tau*. The comparisons of prediction result and experimental data are shown in Figure 3.16B-3.16C.

The data file of solving this problem for SBML-PET is shown in Figure 3.17.

```
# Data file for JAK-STAT model

#************************************************************
#  PAREMETERS TO BE ESTIMATED
#************************************************************
 PARAMETER ID     MINIMUM VALUE    MAXIMUM VALUE
    v1                  1e-2              1e5
    r1                  1e-4              1e4
    r3                  1e-4              1e4
    r4                  1e-4              1e4
    tau                  1                15

#************************************************************
#  Experimental Data
#************************************************************
TOTAL NUMBER of EXPERIMENTAL CONDITIONS = 1
# The number of Time Courses at each Experimental Condition
AT CONDITION 1, The TOTAL NUMBER of TIME COURSES is 1
DATA OF TIME COURSE 1 AT CONDITION 1
NUM_SAMPLING_POINTS 16 NUM_EXP_DATA 2
time          v2+2*v3          SD          v1+v2+2*v3    SD
0             0                0.0768          1          0.084
2             0.34401          0.0519          0.9275     0.046
 …            …                …               …          …
60           0.2650           0.0737          0.971      0.082
Normalized        1                              1

#************************************************************
#  Constraints Information
#************************************************************
# The number of Constraints at each Experimental Condition
AT CONDITION 1, The TOTAL NUMBER of CONSTRAINTS is 0
```

Figure 3.17 Parameter estimation data file for JAK-STAT model in SBML-PET

3.5.4 Other examples

Three other tested examples covering different cases of parameter estimation are provided in the manual document (http://sysbio.molgen.mpg.de/SBML-PET/Manual_SBML-PET.pdf). A yeast aging model demonstrates how SBML-PET supports events. Real experimental data with mathematical expression are used to estimate parameters in the model of irreversible inhibition of HIV proteinase (Mendes and Kell, 1998). Parameter estimation for different experimental conditions (different extra-cellular trypophan concentrations) are shown in the dynamic model of E. Coli tryptophan operon (Bhartiya et al., 2003).

The computation time necessary to find the best fit solution depends on the complexity of the ODE system (the number of ODE and parameters), the number of experimental data and the CPU speed of the computer. For simple problems (with a few number of species and data), it takes several minutes and hours to get the best result. For complex problem (with a lot of species or data), it takes about one day to get the best result. The time of finding best solution for the tested examples shown in manual file is about 15 minutes for model of Michaelis-Menten Equations, Yeast Aging model and JAK-STAT model, 5 hours for HIV Model and 2 hours for dynamic model of *E. Coli* tryptophan operon. The examples were run in Windows 2000 with CPU of AMD Athlon 64 3200, 2.2 GHz.

Chapter 4 Constrain-based Modeling and Kinetic Analysis of Smad Dependent TGF-β Signaling Pathway

4.1 Introduction

The transforming growth factor β (TGF-β) superfamily consists of TGF-βs, bone morphogenetic proteins (BMPs), activins and related proteins. These proteins regulate numerous cellular processes, such as cell proliferation, differentiation, apoptosis and specification of developmental fate (Feng and Derynck, 2005; Shi and Massague, 2003). TGF-β initiates signaling by forming a ligand-receptor complex with the type I and type II receptors at cell surface. The activated receptor complex propagates the signal inside by phosphorylating the receptor-regulated Smad (R-Smad). Activated R-Smad then forms a heteromeric complex with common mediated Smad (Co-Smad), Smad4. These complexes accumulate in the nucleus and regulate gene expression in a cell-type-specific manner through interactions with transcription factors, coactivators and corepressors (Massague et al., 2005). The nuclear Smad complexes are then dephosphorylated by Smad phosphatase (Lin et al., 2006). Another important player is inhibitory Smad (I-Smad), Smad7, which recruits Smurf to the TGF-β receptor complex to facilitate the ubiquitin dependent degradation of receptors (Ogunjimi et al., 2005). Both R-Smad and Smad4 continuously shuttle between the cytoplasm and nucleus in uninduced cells and also in the presence of TGF-β signal (Pierreux et al., 2000; Schmierer and Hill, 2005; Xu et al., 2002). The receptors and activated ligand-receptor complex are internalized through two distinct endocytic routes, the clathrin-dependent endocytosis and the caveolar lipid-raft mediated endocytosis (Di Guglielmo et al., 2003; Mitchell et al., 2004). Figure 4.1 schematically depicts the Smad dependent TGF-β signaling pathways.

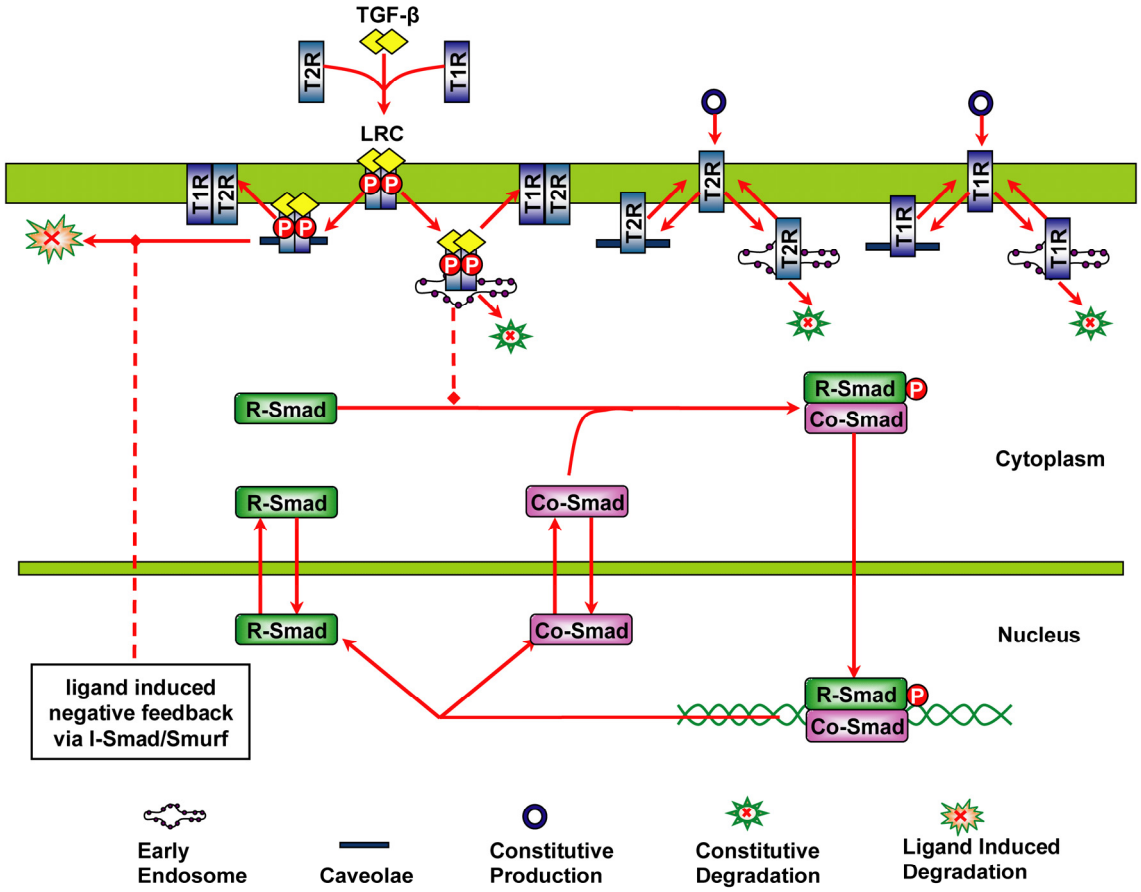

Figure 4.1 Schematic representation of Smad dependent TGF-β signaling pathway

Quantitative modeling studies of signaling pathways have been successfully applied to understand complex cellular processes (Hoffmann et al., 2002; Klipp et al., 2005b; Schoeberl et al., 2002; Swameye et al., 2003). Most of the quantitative models are validated by fitting western-blot data. Due to the complexity of the model and the quality of the data, over-fitting is a common problem for parameter estimation. When the model has many estimated parameters and the corresponding data is a few, the over-fitting problem might lead to some unwarranted conclusions because the parameter set in the model might be a special domain of possible parameter sets which can similarly fit the data well, but result in different predictions. On the other hand, genome-scale constraint-based models of metabolism have been constructed and used to successfully interpret and predict cellular behavior (Price et al., 2003). Constraint-based modeling is an effective method to narrow the range of the possible parameter space for quantitative models. However, there is less attention to apply the constraint-based modeling for the quantitative analysis of signaling pathways. Here, we proposed a constraint-based modeling method to build a comprehensive mathematical model

for the Smad-dependent TGF-β signaling pathway by fitting experimental data and incorporating the qualitative constraints from the experimental analysis.

So far, several mathematical models have been proposed for the TGF-β signaling pathways. Vilar et al. proposed a concise model for the TGF-β receptor trafficking network (Vilar et al., 2006). On the other hand, Clarke et al. (Clarke et al., 2006) and Melke et al. (Melke et al., 2006) proposed the models for the Smad nucleocytoplasmic shuttling and Smad phosphorylation response to the TGF-β signal without considering the receptor trafficking steps. In this study, we established a comprehensive mathematical model for the TGF-β signaling pathway, which includes the signal transduction, the receptor endocytosis, the Smad nucleocytoplasmic shuttling, and the ligand induced negative feedback. The simulation analysis of the model agrees well with the experimental analysis of TGF-β pathway, such as the time course of nuclear phosphorylated Smad, the subcellular location of Smad, and Smad phosphorylation response to different concentrations of TGF-β. We employed the mathematical model to study the dynamic relationship between receptor trafficking at the cell surface and the activation of the phosphorylated Smad in the nucleus. The simulation results indicate that the signal response to TGF-β is regulated by the balance between clathrin dependent endocytosis and non-clathrin mediated endocytosis.

4.2 Materials and Methods

4.2.1 Mathematical model of the Smad dependent TGF-β signaling pathway

We proposed here a comprehensive model for Smad dependent TGF-β signaling pathway in mammalian cells, which includes three modules: receptor trafficking and signaling; Smad nucleocytoplasmic shuttling and I-Smad negative regulation. In the module of receptor trafficking, the model takes into account the following essential elements:

- constitutive receptor synthesis and degradation;
- receptor and ligand-receptor complex trafficking by two distinct endocytic routes;
- the distribution of receptors and ligand-receptor complex in different pools, such as membrane surface, early endosome and caveolar lipid-raft;
- the formation of activated receptor complex induced by TGF-β.

In the module of Smad nucleocytoplasmic shuttling, the signal events we consider are:

- Smad nucleocytoplasmic shuttling;
- Smad complex formation;
- nuclear Smad complex dephosphorylation by nuclear phosphatase.

In the module of I-Smad negative regulation, the negative feedback contributing to the ligand-induced degradation of the receptors by I-Smad is simplified and modeled as a black box.

In order to model this complex signaling pathway, we made the following assumptions for the model:

(1) Experimental analyses indicate that TGF-β receptor trafficking is not affected by TGF-β stimulation (Di Guglielmo et al., 2003; Mitchell et al., 2004). We assume that the corresponding internalization and recycling rates for type I receptor, type II receptor and activated ligand-receptors are the same.

(2) Referring to the concise model of signal processing in the TGF-β ligand-receptor network (Vilar et al., 2006), we assume that the rate of sequential formation of ligand-receptor complex is proportional to the amount of ligand, type I receptor and type II receptor.

(3) The experimental data shown by Lin, *et al.* indicate that the variation of total amount of Smad is small (Lin et al., 2006). We assume that the total amount of Smad is constant. Therefore, the production and degradation of Smad are not considered in this model.

(4) Previous experimental analyses suggest that I-Smad/Smurf complex targets the lipid-raft-bound receptor for degradation which leads to the ligand-induced negative feedback (Di Guglielmo et al., 2003; Ogunjimi et al., 2005). We assume that the rate of ligand-induced degradation of receptors is proportional to the amount of nuclear phoshporylated Smad and the receptor complex in caveolar lipid-raft.

In order to compare with the published data, we used Smad2 denoting R-Smad. We used the law of mass-action to describe the rate of signal transduction steps. The time-dependent

changes of the concentrations of the signaling proteins and protein complexes are determined by the following system of ordinary differential equations:

$$\frac{d[T1R_{Surf}]}{dt} = v_{T1R} - ki_{Cave}[T1R_{Surf}] + kr_{Cave}[T1R_{Cave}] + kr_{EE}[LRC_{EE}] - ki_{EE}[T1R_{Surf}] + kr_{EE}[T1R_{EE}] + kr_{Cave}[LRC_{Cave}] - k_{LRC}[TGF\beta][T1R_{Surf}][T2R_{Surf}] \tag{4.1}$$

$$\frac{d[T1R_{Cave}]}{dt} = ki_{Cave}[T1R_{Surf}] - kr_{Cave}[T1R_{Cave}] \tag{4.2}$$

$$\frac{d[T1R_{EE}]}{dt} = ki_{EE}[T1R_{Surf}] - kr_{EE}[T1R_{EE}] - k_{deg}^{T1R}[T1R_{EE}] \tag{4.3}$$

$$\frac{d[T2R_{Surf}]}{dt} = v_{T2R} - ki_{Cave}[T2R_{Surf}] + kr_{Cave}[T2R_{Cave}] + kr_{EE}[LRC_{EE}] - ki_{EE}[T2R_{Surf}] + kr_{EE}[T2R_{EE}] + kr_{Cave}[LRC_{Cave}] - k_{LRC}[TGF\beta][T1R_{Surf}][T2R_{Surf}] \tag{4.4}$$

$$\frac{d[T2R_{Cave}]}{dt} = ki_{Cave}[T2R_{Surf}] - kr_{Cave}[T2R_{Cave}] \tag{4.5}$$

$$\frac{d[T2R_{EE}]}{dt} = ki_{EE}[T2R_{Surf}] - kr_{EE}[T2R_{EE}] - k_{deg}^{T2R}[T2R_{EE}] \tag{4.6}$$

$$\frac{d[LRC_{Surf}]}{dt} = k_{LRC}[TGF\beta][T1R_{Surf}][T2R_{Surf}] - ki_{Cave}[LRC_{Surf}] - ki_{EE}[LRC_{Surf}] \tag{4.7}$$

$$\frac{d[LRC_{Cave}]}{dt} = ki_{Cave}[LRC_{Surf}] - kr_{Cave}[LRC_{Cave}] - k_{lid}[LRC_{Cave}][Smads\_Complex_{nuc}] \tag{4.8}$$

$$\frac{d[LRC_{EE}]}{dt} = ki_{EE}[LRC_{Surf}] - kr_{EE}[LRC_{EE}] - k_{cd}[LRC_{EE}] \tag{4.9}$$

$$\frac{d[Smad2_{cyt}]}{dt} = -k_{imp}^{Smad2}[Smad2_{cyt}] + \frac{k_{exp}^{Smad2}[Smad2_{nuc}]V_{nuc}}{V_{cyt}} - k_{Smads\_Complex}[Smad2_{cyt}][Smad4_{cyt}][LRC_{EE}] \tag{4.10}$$

$$\frac{d[Smad2_{nuc}]}{dt} = \frac{k_{imp}^{Smad2}[Smad2_{cyt}]V_{cyt}}{V_{nuc}} - k_{exp}^{Smad2}[Smad2_{nuc}] + k_{diss}^{Smads\_Complex}[Smads\_Complex_{nuc}] \quad (4.11)$$

$$\frac{d[Smad4_{cyt}]}{dt} = -k_{imp}^{Smad4}[Smad4_{cyt}] + \frac{k_{exp}^{Smad4}[Smad4_{nuc}]V_{nuc}}{V_{cyt}} - k_{Smads\_Complex}[Smad2_{cyt}][Smad4_{cyt}][LRC_{EE}] \quad (4.12)$$

$$\frac{d[Smad4_{nuc}]}{dt} = \frac{k_{imp}^{Smad4}[Smad4_{cyt}]V_{cyt}}{V_{nuc}} - k_{exp}^{Smad4}[Smad4_{nuc}] + k_{diss}^{Smads\_Complex}[Smads\_Complex_{nuc}] \quad (4.13)$$

$$\frac{d[Smads\_Complex_{cyt}]}{dt} = k_{Smads\_Complex}[Smad2_{cyt}][Smad4_{cyt}][LRC_{EE}] - k_{imp}^{Smads\_Complex}[Smads\_Complex_{cyt}] \quad (4.14)$$

$$\frac{d[Smads\_Complex_{nuc}]}{dt} = \frac{k_{imp}^{Smads\_Complex}[Smads\_Complex_{cyt}]V_{cyt}}{V_{nuc}} - k_{diss}^{Smads\_Complex}[Smads\_Complex_{nuc}] \quad (4.15)$$

$$\frac{d[TGF\beta]}{dt} = (kr_{Cave}[LRC_{Cave}] + kr_{EE}[LRC_{EE}] - k_{LRC}[TGF\beta][T1R_{surf}][T2R_{surf}])V_{cyt}/V_{extra} \quad (4.16)$$

The model is composed of an ordinary differential equations system with 16 state variables and 20 parameters. The values and the corresponding biological meaning of the parameters are listed in Table 4.1. The initial conditions and the biological meaning of the state variables are listed in Table 4.2.

Table 4.1 Parameter values in the TGF-β model

Parameter	Biological meaning	Value	Unit
v_{T1R}	type I receptor production rate constant	0.0103	nM/min
v_{T2R}	type II receptor production rate constant	0.02869	nM/min
ki_{EE}	internalization rate constant of receptor to early endosome	0.33	min^{-1}
kr_{EE}	recycling rate constant of receptor from early endosome to cell surface	0.033	min^{-1}
ki_{Cave}	internalization rate constant of receptor to caveolar lipid-raft	0.33	min^{-1}
kr_{Cave}	recycling rate constant of receptor from caveolar lipid-raft	0.03742	min^{-1}
k_{cd}	degradation rate constant for ligand-receptor complex in early endosome	0.005	min^{-1}
k_{LRC}	ligand-receptor complex formation rate constant	2197	$nM^{-2}min^{-1}$
k_{lid}	ligand induced degradation rate constant for ligand-receptor complex	0.02609	min^{-1}
k_{deg}^{T1R}	degradation rate constant for type I receptor in early endosome	0.005	min^{-1}
k_{deg}^{T2R}	degradation rate constant for type II receptor in early endosome	0.025	min^{-1}
k_{imp}^{Smad2}	nuclear import rate constant for Smad2	0.16	min^{-1}
k_{exp}^{Smad2}	nuclear export rate constant for Smad2	1	min^{-1}
k_{imp}^{Smad4}	nuclear import rate constant for Smad4	0.08	min^{-1}
k_{exp}^{Smad4}	nuclear export rate constant for Smad4	0.5	min^{-1}
$k_{Smads\_Complex}$	formation rate constant for the phosphorylated Smad complex	6.85×10^{-5}	$nM^{-2}min^{-1}$
$k_{imp}^{Smads\_Complex}$	nuclear import rate constant for the phosphorylated Smad complex	0.16	min^{-1}
$k_{diss}^{Smads\_Complex}$	dissociation rate constant for the nuclear phosphorylated Smad complex	0.1174	min^{-1}

Parameter	Biological meaning	Value	Unit
V_{cyt}/V_{nuc}	ratio of cytoplasmic to nuclear volume	3.0	
V_{cyt}/V_{extra}	ratio of cytoplasmic volume to the average extracellular medium volume per cell	0.001	

Table 4.2 Initial conditions of state variables (proteins) in the model

Variable symbol	Biological meaning	Value of the initial condition
$T1R_{Surf}$	type I receptor at cell surface	0.237 nM
$T1R_{Cave}$	type I receptor in caveolar lipid-raft	2.092 nM
$T1R_{EE}$	type I receptor in early endosome	2.06 nM
$T2R_{Surf}$	type II receptor at cell surface	0.202 nM
$T2R_{Cave}$	type II receptor in caveolar lipid-raft	1.778 nM
$T2R_{EE}$	type II receptor in early endosome	1.148 nM
LRC_{Surf}	ligand-receptor complex at cell surface	0 nM
LRC_{Cave}	ligand-receptor complex in caveolar lipid-raft	0 nM
LRC_{EE}	ligand-receptor complex in early endosome	0 nM
$Smad2_{cyt}$	Smad2 in the cytoplasm	492.61 nM
$Smad2_{nuc}$	Smad2 in the nucleus	236.45 nM
$Smad4_{cyt}$	Smad4 in the cytoplasm	1149.4 nM
$Smad4_{nuc}$	Smad4 in the nucleus	551.72 nM
$Smads\_Complex_{cyt}$	phosphorylated Smad complex in the cytoplasm	0 nM
$Smads\_Complex_{nuc}$	phosphorylated Smad complex in the nucleus	0 nM
$TGF\beta$	TGF-β in the extracellular medium	0.08 nM for 2 ng/ml

The initial conditions of the variables are set as the steady state concentration of the corresponding proteins in the uninduced cell which are derived from the parameter values described in Table 4.1.

4.2.2 Derivation of the parameter values

In order to keep the consistency of the parameter values, we derived the parameter values from experimental analysis in epithelial cells. In particularly, most of the data used in this study are based on the experimental analysis of HaCaT (Human keratinocyte cell line) cells. The rational guided the derivation of relative parameter values is described in the following:

- **Parameters for the volume ratio of different compartments:** The model distinguishes the different compartments of extracellular medium, cytoplasm and nucleus where the signaling steps take place. The average cytoplasimc/nuclear volume ratio ($\frac{V_{cyt}}{V_{nuc}}$) was measured as about 3 in reference (Schmierer and Hill, 2005). The volume of HaCaT cell is estimated as about 1.4×10^{-12} Liter (Savini et al., 2000; Watt et al., 1988), therefore, we can derive the volume of cytoplasm and nucleus are 1.05×10^{-12} Liter and 3.5×10^{-13} Liter, respectively. A typical cell culture experiment would have a cell density of 1.0×10^{6} cells/ml. The average extracellular medium volume per cell (V_{extra}) is about 1.0×10^{-9} Liter. Hence, the volume ratio of cytoplasm to extracellular medium ($\frac{V_{cyt}}{V_{extra}}$) is set to a value of 0.001.

- **Rate constant of receptor constitutive degradation**: Pervious experimental data indicate that the type II receptor has a half-life of about 1 hr, whereas type I receptors are more stable, with a half-life of about 4–6 hr (Kavsak et al., 2000). The receptors are constitutively degraded from early endosomes and about 50% percent of initial labeled receptors locate in early endosome (Di Guglielmo et al., 2003). Rescaling the degradation rate of the total receptors to the early endosome receptors results in constitutive degradation rate constant of type I receptor (k_{deg}^{T1R}) and type II receptor (k_{deg}^{T2R}) with value of 0.005 min^{-1} ($\frac{-\ln 0.5}{300\,\text{min}\times 0.5}$) and 0.025 min^{-1} ($\frac{-\ln 0.5}{60\,\text{min}\times 0.5}$), respectively. On the other hand, Figure 3 of reference (Di Guglielmo et al., 2003) shows that only 30% of the initial labeled receptor complex remain in the cell after 8h when the caveolar endocytosis is inhibited. Based on this information, we can derive the

constitutive degradation rate constant of ligand-receptor complex (k_{cd}) is 0.005 min^{-1} ($\frac{-\ln 0.3}{480\,\text{min}\times 0.5}$).

- **Rate constant of receptor internalization and recycling**: Vilar *et al.* derived the rate constant of receptor internalization and recycling based on the experimental observations (Vilar et al., 2006). On the other hand, reference (Di Guglielmo et al., 2003) indicate that receptors are internalized through clathrin and non-clathrin independent endocytosis route with similar rates. Therefore, we choose the internalization rate constant of receptors (ki_{EE} and ki_{Cave}) with the value of 0.33 min^{-1} which is estimated in reference (Vilar et al., 2006). The recycling rate constant (kr_{EE}) for the receptors recycled from early endosome back to the cell surface is set to a value of 0.033 min^{-1}, which is derived in reference (Vilar et al., 2006).

- **Rate constant for Smad nuclear import and export**: The nuclear import rate of Smad2 (k_{imp}^{Smad2}) was experimentally measured with the value of 0.16 min^{-1} in reference (Schmierer and Hill, 2005). Reference (Schmierer and Hill, 2005) also measured that the ratio of mean nuclear fluorescence to mean cytoplasmic fluorescence (R) of Smad2 in uninduced cells is about 0.5. The R ratio of Smad2 is equivalent to the ratio of nuclear to cytoplasmic Smad2 concentrations. We can derive $R = \frac{k_{imp}^{Smad2} \cdot V_{cyt}}{k_{exp}^{Smad2} \cdot V_{nuc}}$ according to equation (4.31). Therefore, the value for the nuclear export rate constant (k_{exp}^{Smad2}) in our model is about 1 min^{-1} ($k_{exp}^{Smad2} = \frac{k_{imp}^{Smad2}}{R} \cdot \frac{V_{cyt}}{V_{nuc}}$). On the other hand, Figure 5A of reference (Schmierer and Hill, 2005) also shows that the nuclear export rate of Smad4 is about half of the nuclear export rate of Smad2. Hence, the nuclear export rate constant of Smad4 (k_{exp}^{Smad4}) is 0.5 min^{-1}. Since the R ratio of Smad4 is measured with a value of about 0.5 (Schmierer and Hill, 2005), we can derive the nuclear import rate constant of Smad4 (k_{imp}^{Smad4}) a value of 0.08 min^{-1}.

- **Nuclear import rate constant of phosphorylated Smad:** Figure 3 of reference (Xu et al., 2000) shows that nuclear import rate between phoshporylated and unphosphorylated Smad2 is similar. Therefore, we set the import rate constant for the phosphorylated

Smad complex ($k_{imp}^{Smads\_Complex}$) to the same value as the import rate constant for unphosphorylated Smad2, which is 0.16 min^{-1}.

There are 7 unknown parameter values in the model that are required to be estimated, i.e. v_{T1R}, v_{T2R}, kr_{Cave}, k_{LRC}, k_{lid}, $k_{Smads\_Complex}$ and $k_{diss}^{Smads\_Complex}$.

4.2.3 Steady state analysis of the model for the uninduced cell

We performed steady state analysis of the model for the uninduced cell without TGF-β stimulation. When there is no TGF-β present to the cells, the concentrations of the ligand-receptor complex and phosphorylated Smad complex are assumed to be zero.

For type I receptor, we can derive the following algebraic equations for the steady state of the unindcued cell:

$$\frac{d[T1R_{Surf}]_{ss}}{dt} = v_{T1R} - ki_{Cave}[T1R_{Surf}]_{ss} + kr_{Cave}[T1R_{Cave}]_{ss} - ki_{EE}[T1R_{Surf}]_{ss} + kr_{EE}[T1R_{EE}]_{ss} = 0 \tag{4.17}$$

$$\frac{d[T1R_{Cave}]_{ss}}{dt} = ki_{Cave}[T1R_{Surf}]_{ss} - kr_{Cave}[T1R_{Cave}]_{ss} = 0 \tag{4.18}$$

$$\frac{d[T1R_{EE}]_{ss}}{dt} = ki_{EE}[T1R_{Surf}]_{ss} - kr_{EE}[T1R_{EE}]_{ss} - k_{deg}^{T1R}[T1R_{EE}]_{ss} = 0 \tag{4.19}$$

The steady state concentrations of type I receptor at cell surface, in the lipid-raft and in the early endosome, obtained by solving the systems of algebraic equations (4.17-4.19), are

$$[T1R_{Surf}]_{ss} = \frac{v_{T1R}(kr_{EE} + k_{deg}^{T1R})}{k_{deg}^{T1R} \cdot ki_{EE}} \tag{4.20}$$

$$[T1R_{Cave}]_{ss} = \frac{v_{T1R}(kr_{EE} + k_{deg}^{T1R})ki_{Cave}}{k_{deg}^{T1R} \cdot ki_{EE} \cdot kr_{Cave}} \tag{4.21}$$

$$[T1R_{EE}]_{ss} = \frac{v_{T1R}}{k_{deg}^{T1R}} \tag{4.22}$$

The steady state concentrations of type II receptor at cell surface, in the lipid-raft and in the early endosome are:

$$[T2R_{Surf}]_{ss} = \frac{v_{T2R}(kr_{EE} + k_{deg}^{T2R})}{k_{deg}^{T2R} \cdot ki_{EE}} \tag{4.23}$$

$$[T2R_{Surf}]_{ss} = \frac{v_{T2R}(kr_{EE} + k_{deg}^{T2R})}{k_{deg}^{T2R} \cdot ki_{EE}} \tag{4.24}$$

$$[T2R_{EE}]_{ss} = \frac{v_{T2R}}{k_{deg}^{T2R}} \tag{4.25}$$

For the Smad2 in the cytoplasm and nucleus, the following algebraic equations should be satisfied for the steady state of the unindcued cell:

$$\frac{d[Smad2_{cyt}]_{ss}}{dt} = -k_{imp}^{Smad2}[Smad2_{cyt}]_{ss} + \frac{k_{exp}^{Smad2}[Smad2_{nuc}]_{ss}V_{nuc}}{V_{cyt}} = 0 \tag{4.26}$$

$$\frac{d[Smad2_{nuc}]_{ss}}{dt} = \frac{k_{imp}^{Smad2}[Smad2_{cyt}]_{ss}V_{cyt}}{V_{nuc}} - k_{exp}^{Smad2}[Smad2_{nuc}]_{ss} = 0 \tag{4.27}$$

Because the total amount of Smad2 is a constant, mass conservation reads:

$$N_{Smad\,2}^{total} = V_{cyt}[Smad2_{cyt}]_{ss} + V_{nuc}[Smad2_{nuc}]_{ss} \tag{4.28}$$

Solving the algebraic equations (4.26-4.28) yields the steady state concentrations of cytoplasmic and nuclear Smad2 in the uninduced cell as following:

$$[Smad2_{cyt}]_{ss} = \frac{N_{Smad2}^{total} \cdot k_{exp}^{Smad2}}{V_{cyt}(k_{imp}^{Smad2} + k_{exp}^{Smad2})} \tag{4.29}$$

$$[Smad2_{nuc}]_{ss} = \frac{N_{Smad2}^{total} \cdot k_{imp}^{Smad2}}{V_{nuc}(k_{imp}^{Smad2} + k_{exp}^{Smad2})} \tag{4.30}$$

The ratio of nuclear to cytoplasmic Smad2 concentration corresponds to the ratio of mean nuclear fluorescence to mean cytoplasmic fluorescence (R) of Smad2, according to equations (29-30), which is

$$R = \frac{[Smad2_{nuc}]_{ss}}{[Smad2_{cyt}]_{ss}} = \frac{k_{imp}^{Smad2} \cdot V_{cyt}}{k_{exp}^{Smad2} \cdot V_{nuc}} \tag{4.31}$$

The steady state concentrations of cytoplasmic and nuclear Smad4 in the uninduced cell are

$$[Smad4_{cyt}]_{ss} = \frac{N_{Smad4}^{total} \cdot k_{exp}^{Smad4}}{V_{cyt}(k_{imp}^{Smad4} + k_{exp}^{Smad4})} \tag{4.32}$$

$$[Smad4_{nuc}]_{ss} = \frac{N_{Smad4}^{total} \cdot k_{imp}^{Smad4}}{V_{nuc}(k_{imp}^{Smad4} + k_{exp}^{Smad4})} \tag{4.33}$$

The total amount of Smad2 (N_{Smad2}^{total}) and Smad4 (N_{Smad4}^{total}) are estimated to be 6×10^{-19} mol per cell and 1.4×10^{-18} mol per cell in HaCaT cells, which correspond to about 3.6×10^{5} and 8.4×10^{5} molecules per cell, respectively (He et al., 2006).

Without the treatment of TGF-β, the concentrations of signaling proteins will arrive at steady state balancing protein production, degradation, receptor endocytosis and Smad protein nucleocytoplasmic shuttling. We set the initial concentrations of the signaling proteins as their steady state values before TGF-β is added.

4.2.4 Parameter estimation by constraint-based modeling

The parameter estimation for the 7 unknown values of rate constants was done using a modified version of the tool SBML-PET (Zi and Klipp, 2006), which incorporates stochastic ranking evolution strategy (SRES) for parameter estimation jobs. SRES is an evolutionary optimization algorithm that uses stochastic ranking as the constraint handling technique and is described in Appendix B (Runarsson and Yao, 2000). The objective of the parameter estimation is to find the most feasible parameters in the model that reproduce the quantitative experimental data for the TGF-β signaling pathway. At the same time, the model with the estimated parameters should satisfy some qualitative experimental observation of this pathway. Therefore, the corresponding quantitative time course data is used in the objective function definition and the qualitative data is coded as constraints during the optimization process. We used two time courses of Smad phosphorylation for parameter estimation:

(1) The quantitative western blot data reported in Figure 1A in Lin *et al.* (Lin et al., 2006) shows the time course of Smad2 in the whole HaCaT cell extract in the presence of continuous TGF-β stimulation for 24 hours. We normalized the data to its maximum value which gives us the dynamic change of the protein. We also normalized corresponding simulation data to its maximum value so that we can compare the experimental data and the simulation data.

(2) The quantitative western blot data reported in Figure 1C in Lin *et al.* (Lin et al., 2006) gives us the time course of Smad2 phosphorylation under a different condition. TGF-β (2 ng/ml) is added at first 30 minutes, then washed out and type I receptor kinase inhibitor SB431542 is added to terminate the signal.

Other qualitative information from the literatures is used as constraints encoded within SRES method.

(1) The number of total receptors per cell falls into the range of 1000 to 100000 per cell according to the distribution of TGF-β receptors in a wide spectrum of cell types reported in Wakefield et al. (Wakefield et al., 1987).

(2) Experimental results of receptor distribution at the cell surface, in the early endosomes and caveolin-positive vesicles indicate that about 40~50% percent of total receptors are located in the early endosomes (Di Guglielmo et al., 2003; Mitchell et al., 2004). We set a constraint for the distribution of receptors, that the about at least 30% and at maximum 60% of receptors locate in the early endosome.

(3) Schmierer et al. quantified the Smad redistribution in HaCaT cells upon 2 ng/ml TGF-β with photobleaching experiments using EGFP-Smad (enhanced green fluorescent protein) fusions (Schmierer and Hill, 2005). The study indicates that the ratio of nuclear to cytoplasmic Smad2 concentrations is about 2.5 after 1 hour TGF-β treatment. After TGF-β stimulus, about 80% of the Smad2 in the nuclues is phosphorylated or complexed. We set constraints for the corresponding Smad2 distribution: the ratio of nuclear to cytoplasmic Smad2 concentrations should be in the range of 1.5~3.5 and 60%~90% of the Smad2 in the nucleus is phosphorylated or complexed after 1 hour TGF-β treatment.

(4) Previous experimental data indicate that 0.5 ng/ml (20 pM) of TGF-β is sufficient to yield a maximal response of Smad phosphorylation after 1 hour TGF-β treatment

(Bakin et al., 2000; Goumans et al., 2002). We encode this information as a constraint that the ratio of Smad phosphorylation level with the dose of 0.5 ng/ml to that with the dose of 2 ng/ml TGF-β at 60th minute should be larger than 0.9 and smaller than 1.1.

4.3 Results and Discussion

4.3.1 Comparison of kinetic simulation with experimental analysis

We first check whether the model obtained by constraint-based modeling method can reproduce the experimental data used for parameter estimation, which can give us the information about quality of "in-sample fit" (how good is the model fitting the data used for parameter estimation). Experiments reported that the signal would peak at about 30-60 min after TGF-β addition. The result shown in Figure 4.2A indicates that the simulated time courses of Smad2 phosphorylation agree with the experimental data (Lin et al., 2006). The model is also able to reproduce the experimental observation that the treatment with type I receptor kinase inhibitor SB431542 will cause rapid decrease of the nuclear phosphorylated Smad2 level (Inman et al., 2002; Lin et al., 2006) (Figure 4.2B). Taken these simulation results together, the model has a good "in-sample fit" for the data.

We next asked whether the model has a good match to other experimental data that were not used for parameter estimation. This test can be regarded as "out-sample fit" or model validation. The result shown in Figure 4.2C indicates that the simulated time courses of nuclear phosphorylated Smad2 agree well with the experimental data (Inman et al., 2002). As a further test, we calculated the subcellular location of Smad2 from the simulation result of the model. The results are in agreement with previous reports that TGF-β causes a change in the overall Smad2 distribution from predominantly cytoplasmic to predominantly nuclear (Inman et al., 2002; Pierreux et al., 2000; Schmierer and Hill, 2005; Xu et al., 2002). After TGF-β treatment, Smad2 proteins are rapidly accumulated in the nucleus and then return to the cytoplasm (Figure 4.2D).

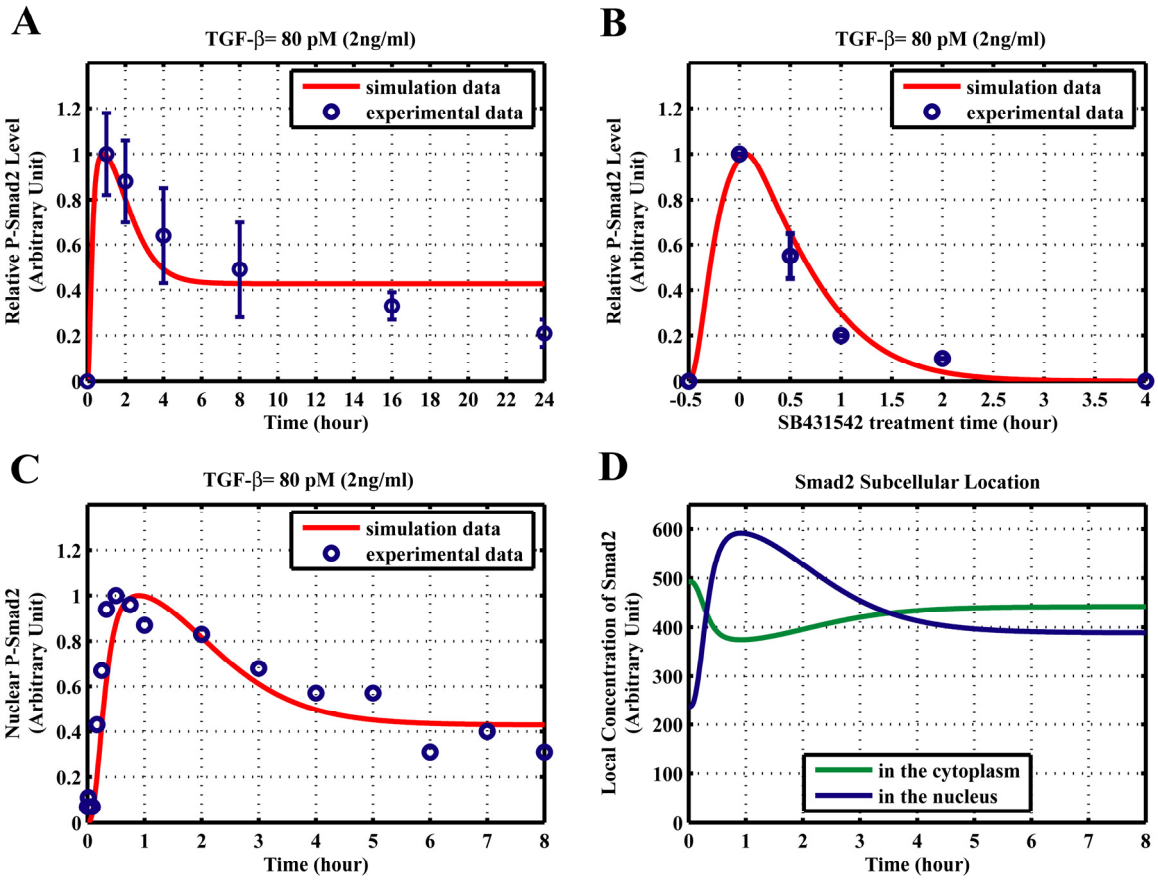

Figure 4.2 Comparison of experimental analysis and simulation results from the model obtained by constraint-based modeling method. (A-B) for "in-sample fit". (C-D) for "out-sample fit". (A) Comparison of the model time course and experimental time course of Smad2 phosphorylation with 24 hours TGF-β treatment. The experimental data is normalized from Figure 1A in Lin et al. (Lin et al., 2006). (B) Effect of type I receptor kinase inhibitor SB431542. Cells were treated with TGF-β for 30 minutes, then were washed out TGF-β at 30th minute and added SB431542 to prevent rephosphorylation of Smad2. The experimental data is normalized from Figure 1C in Lin et al. (Lin et al., 2006). (C) Comparison of the model time course with an experimental time course of nuclear phosphorylated Smad2 after TGF-β treatment (80pM, 2 ng/ml). The western-blot data reported by Inman et al. (Fig.1A, top panel) is quantified with Scion Image software (Inman et al., 2002). (D) Subcellular location of Smad2 after TGF-β treatment (80 pM). The concentrations shown here refer to the local concentrations in cytoplasm and nucleus.

Finally, we tested whether the model can predict well the signal response to different dose of TGF-β. Previous experimental data indicate that the response of Smad phosphorylation

after 1 hour TGF-β treatment will be saturated when the concentration of TGF-β is larger than 0.5 ng/ml (Bakin et al., 2000; Goumans et al., 2002). The model successfully predicts the dose-response of Smad phosphorylation upon different concentrations of TGF-β (Figure 4.3).

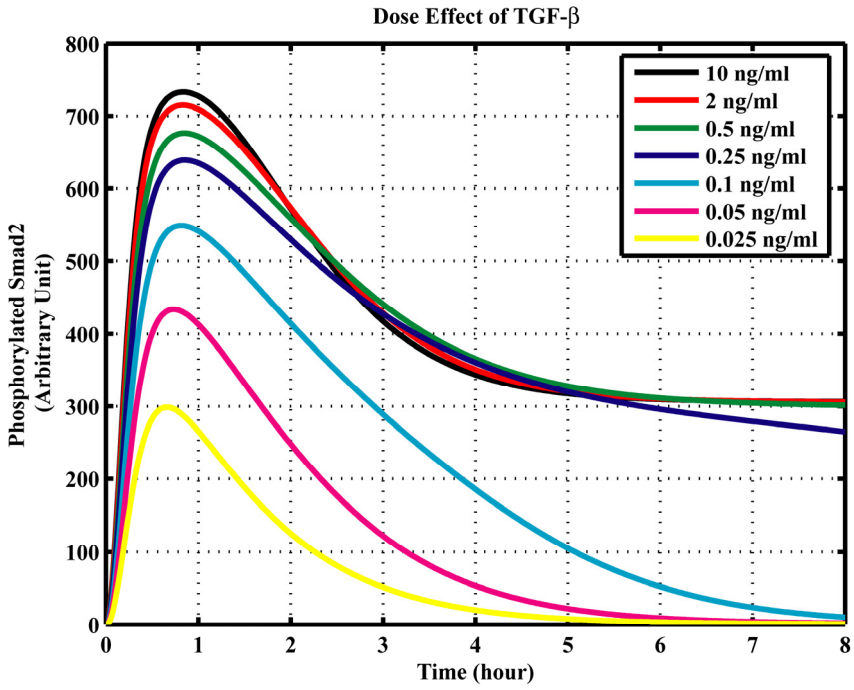

Figure 4.3 Effects on Smad2 phosphorylation by different doses of TGF-β

4.3.2 Model performance is significantly improved by constraint-based modeling

We compared the performance of the model generated by constraint-based modeling method and that of the model obtained by only fitting the time course data. We first compared the experimental data and simulation results from the model obtained by only fitting the time course data. The results shown in Figure 4.4 indicate that the model has been over-fitted to the data used for parameter estimation (Figure 4.4A-4.4B), but it has bad predictions for the data that are not used for parameter estimation. For example, experimental evidence shows that the response of Smad phosphorylation after 1 hour TGF-β treatment will be saturated when the concentration of TGF-β is larger than 0.5 ng/ml (Bakin et al., 2000; Goumans et al., 2002). However, the predicted time course of phosphorylated Smad2 in the model, obtained by only fitting the time course data, is not saturated even for a very high dose of TGF-β (10 ng/ml), which contradicts the experimental results (Figure 4.4C).

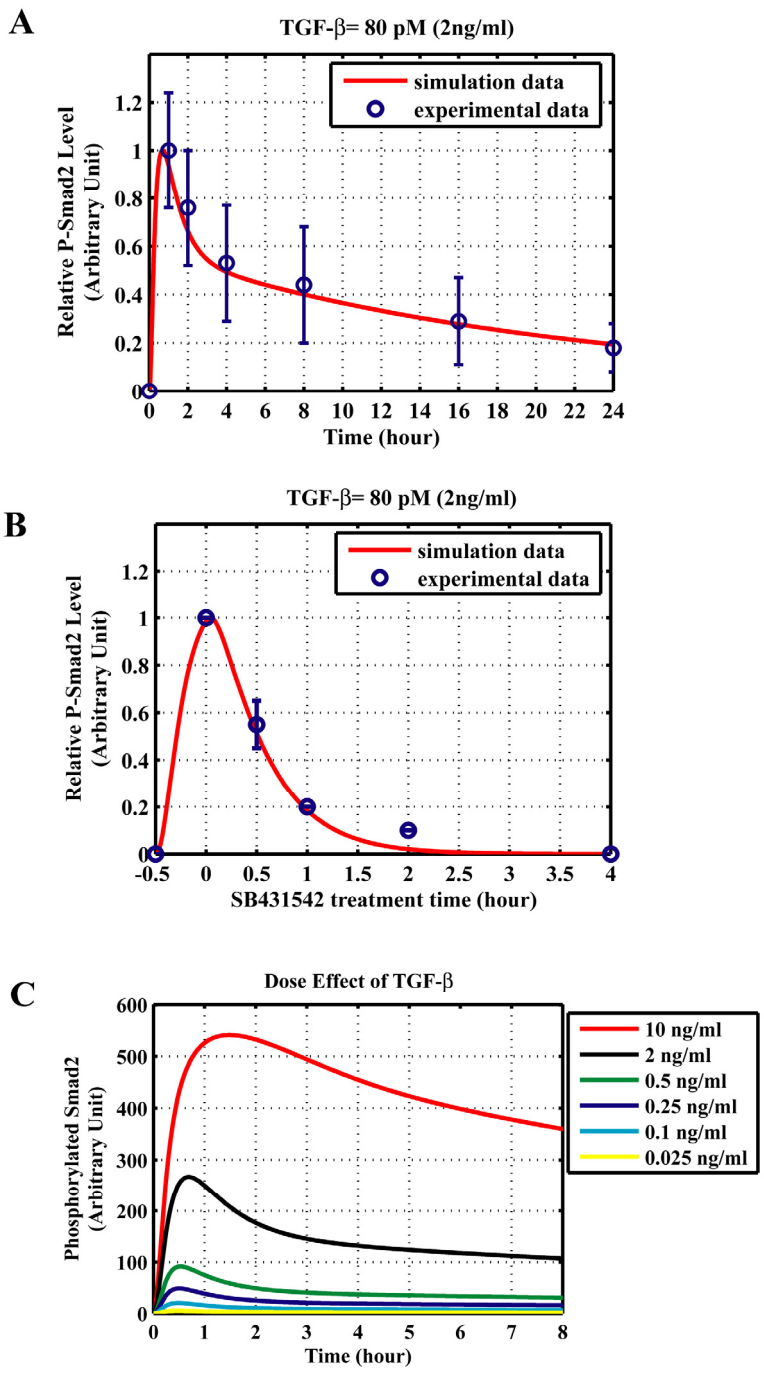

Figure 4.4 Comparison of experimental analysis and simulation results from the model obtained by only fitting the time course data. (A-B) The model has been over-fitted for "in-sample fit". (C) The model has a bad prediction for "out-sample fit". (A) Comparison of the model time course and experimental time course of Smad2 phosphorylation with 24 hours TGF-β treatment. Experimental data is the same as described in Figure 4.2A. (B) Effect of type I receptor kinase inhibitor SB431542. Experimental data is the same as described in Figure 4.2B. (C) Effects on Smad2 phosphorylation by different doses of TGF-β.

On the other hand, we compared the possible variation of parameter sets of the models by these two different methods. For each method, we independently generated 1000 parameter sets which make the model have similar goodness of fitting the time course data of Smad phosphorylation. According to the statistical result for the parameter sets shown in Table 4.3, the ranges of the variation for the 7 estimated parameter values are significantly narrowed by constraint-based modeling method. The 1000 parameter sets data are available online:

http://www.plosone.org/article/info:doi/10.1371/journal.pone.0000936

Table 4.3 Variation range for the estimated parameters in the 1000 parameter sets

	Obtained by constraint-based modeling method			**Obtained by only fitting the time course data**		
Parameter	**Minimum value**	**Maximum value**	**Variation: $\log_{10}$(max/min)**	**Minimum value**	**Maximum value**	**Variation: $\log_{10}$(max/min)**
v_{T1R}	0.007482	0.02712	0.55927	0.001309	0.5195	2.5986
v_{T2R}	0.02167	0.1101	0.70593	0.008435	4.874	2.7618
kr_{Cave}	0.03737	0.04243	0.05515	0.001571	0.04571	1.4638
k_{LRC}	1305	86850	1.8232	0.1088	6.84×10^{6}	7.7981
k_{lid}	0.001011	0.9774	2.9853	1.38×10^{-5}	0.9988	4.8587
$k_{Smads\_Complex}$	2.28×10^{-5}	9.69×10^{-5}	0.628	3.15×10^{-7}	0.000203	2.8089
$k_{diss}^{Smads\_Complex}$	0.09519	0.1298	0.13468	0.07243	0.9985	1.1394

Previous modeling studies of TGF-β signaling pathway indicate that the Smad phosphorylation response to TGF-β is robust to a large variation of some parameters (Clarke et al., 2006; Melke et al., 2006). The large variations in the estimated parameter sets are usually caused by two reasons. The data quality is the main reason for the large variation of parameter sets obtained by only fitting the time course data. When we fit the western-blot data, we actually fit the scaled data. However, the parameter for the scaled coefficient is unknown, which can vary in a large range if we only fitting the scaled time course data without considering other qualitative constraints. Another reason comes from the insensitivity of some parameters to the signal response. For example, the sensitivities of the parameters k_{LRC}, v_{T2R}

and k_{lid} are very small (Figure 4.5), which means the output of the model (Smad phosphorylation level) is robust to the variation of these insensitive parameters.

4.3.3 Sensitivity analysis of the model

We next systematically investigated the sensitivity of all the rate constants and found those whose perturbation the pathway is most sensitive or most robust against. Response sensitivities were used for quantifying the effects of all the rate constants on the concentration of signaling proteins. In this model, we regard nuclear phosphorylated Smad complex (nuclear phosphorylated Smad2) as the readout of the signal in this pathway because experimental results indicate that nuclear phosphorylated Smad complex acts as transcription factor to induce the expression of many genes (Massague et al., 2005; Shi and Massague, 2003). We determined the sensitivity of the integral concentrations of the nuclear phosphorylated Smad complex from the beginning of the TGF-β activation to the end of the simulation time (8 hours). The definition of response sensitivity for nuclear phosphorylated Smad is as following:

$$R_{p_i}^{Smads\_Complex\_n} = \frac{\frac{\partial(Smads\_Complex\_n)}{Smads\_Complex\_n}}{\frac{\partial p_i}{p_i}} \tag{4.34}$$

As shown in Figure 4.5 and Table 4.4, we can divide the parameters into three groups: positive control, negative control and almost no control to the changes of nuclear Smad complex. The strong positive and negative control groups belong to the reactions participating phosphorylation of Smad, receptor endocytosis, type I receptor production. The most negative control on the concentration of nuclear Smad complex is the rate constants corresponding to the dephosphorylation of nuclear Smad complex, which implies that the nuclear phosphatase has a strong negative control on the nuclear phosphorylated Smad level. Finally, we want to point out that all these analyses are based on small perturbations of relative parameters according to the definition of response coefficient.

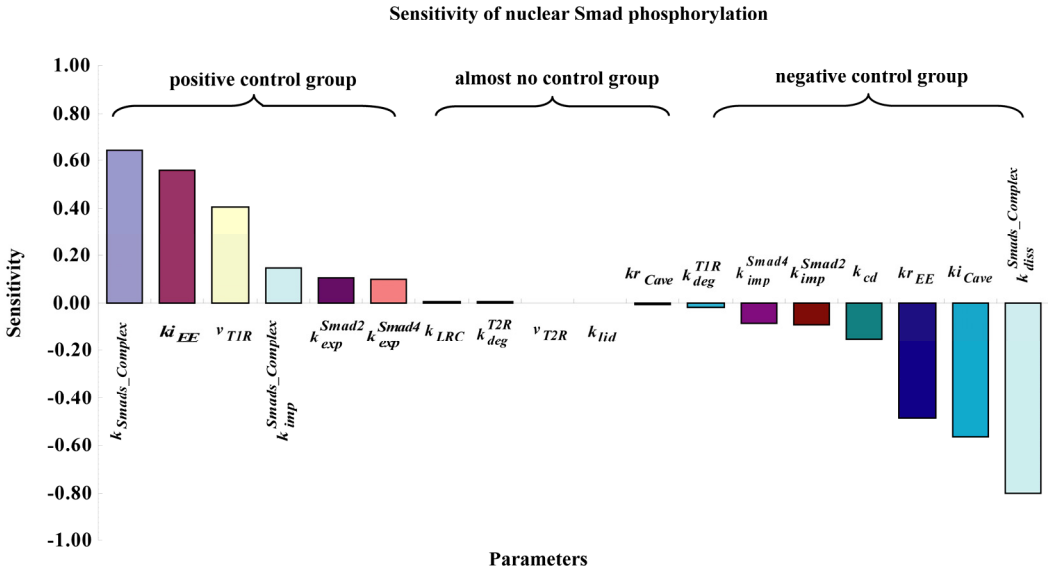

Figure 4.5 Sensitivity analysis result

Table 4.4 Sensitivity analysis result on the rate constant

Rate Constant Symbol	Response Sensitivity	Rate Constant Symbol	Response Sensitivity
$k_{Smads\_Complex}$	0.641	k_{lid}	-4.15×10^{-3}
ki_{EE}	0.561	kr_{Cave}	-1.03×10^{-2}
v_{T1R}	0.402	k_{deg}^{T1R}	-2.34×10^{-2}
$k_{imp}^{Smads\_Complex}$	0.144	k_{imp}^{Smad4}	-8.88×10^{-2}
k_{exp}^{Smad2}	0.104	k_{imp}^{Smad2}	-9.04×10^{-2}
k_{exp}^{Smad4}	9.92×10^{-2}	k_{cd}	-0.150
k_{LRC}	3.80×10^{-3}	kr_{EE}	-0.484
k_{deg}^{T2R}	1.66×10^{-3}	ki_{Cave}	-0.563
v_{T2R}	-2.52×10^{-3}	$k_{diss}^{Smads\_Complex}$	-0.803

The response sensitivities were calculated by numerical determination of the integral response to a change of the rate constants by 0.1%. We take the average of the forward sensitivity (computed by increasing p_i to $1.001 \times p_i$) and backward sensitivity (computed by deceasing p_i to $0.999 \times p_i$) as the response sensitivity for corresponding rate constant. Using relative changes of rate constants less than 0.1% does not lead to a significant improvement of the precision of the values.

4.3.4 The regulation of the signal: balance between clathrin dependent endocytosis and non-clathrin mediated endocytosis

In the experimental studies, potassium depletion can be used to interfere with clathrin-dependent trafficking of receptors, which can inhibit the TGF-β signal (Di Guglielmo et al., 2003; Hayes et al., 2002). On the other hand, nystatin treatment causes the inhibition of the non-clathrin endocytosis pathway (Di Guglielmo et al., 2003). The simulation analysis of the model indicates that the inhibition of clathrin dependent endocytosis causes a transient response of Smad2 phosphorylation (Figure 4.6A). The simulation result also shows that inhibition of non-clathrin dependent endocytosis increases TGF-β signal amplitude, which produces a prolonged signal (Figure 4.6C).

What will happen if both clathrin dependent and non-clathrin dependent endocytosis are inhibited? The simulation results indicate that the key quantity is the ratio of clathrin to non-clathrin dependent endocytosis rate. A transient response of Smad2 phosphorylation appears upon the combination of a strong inhibition of clathrin dependent endocytosis and a weak inhibition of non-clathrin mediated endocytosis, which corresponds to a low ratio of clathrin to non-clathrin dependent endocytosis rate (Figure 4.6D). Furthermore, when both clathrin dependent endocytosis and non-clathrin dependent endocytosis are equally inhibited (a medium ratio of clathrin to non-clathrin dependent endocytosis rate), the signal response of Smad2 phosphorylation is similar as the control (Figure 4.6E). Finally, a prolonged signal response of Smad2 phosphorylation is observed (Figure 4.6F) with the combination of a weak inhibition of clathrin dependent endocytosis and a strong non-clathrin dependent endocytosis (a high ratio of clathrin to non-clathrin dependent endocytosis rate). Therefore, the TGF-β signal response is regulated by the balance between the strength of signal initiation from clathrin dependent endocytosis and the strength of negative feedback in the venue of non-clathrin mediated endocytosis.

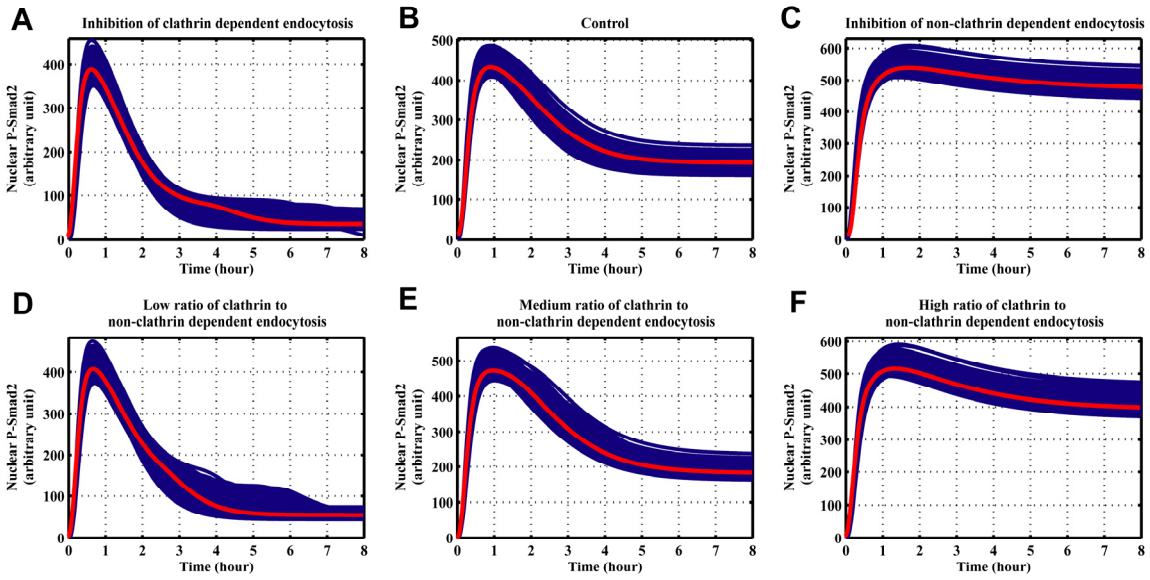

Figure 4.6 Computational simulations of the time course of nuclear phosphorylated Smad2 by the inhibition of different receptor endocytosis in 1000 parameter sets estimated by constraint-based modeling method. The red lines refer the simulations for the parameter values listed in Table 4.1. Blue lines correspond to the 1000 parameter sets. (A) Same parameter values as those in parameter sets with the exception that clathrin dependent internalization rate constant is decreased by a factor of 10: ki_{EE} = 0.033 min^{-1}. (B) Same parameters values as those in parameter sets. (C) Same parameter values as that in parameter sets with the exception that non-clathrin dependent internalization rate constant is decreased by a factor of 10: ki_{Cave} = 0.033 min^{-1}. (D) Same parameter values as those in parameter sets with the exception that ki_{EE} is decreased by a factor of 10 and ki_{Cave} is decreased by a factor of 2: ki_{EE} = 0.033 min^{-1}, ki_{Cave} = 0.165 min^{-1}. (E) Same parameter values as those in parameter sets with the exception that ki_{EE} and ki_{Cave} are decreased by a factor of 10: ki_{EE} = 0.033 min^{-1}, ki_{Cave} = 0.033 min^{-1}. (F) Same parameter values as those in parameter sets with the exception that ki_{EE} is decreased by a factor of 2 and ki_{Cave} is decreased by a factor of 10: ki_{EE} = 0.165 min^{-1}, ki_{Cave} = 0.033 min^{-1}.

Recently Vilar et al. proposed a concise computational model of signal processing in TGF-β superfamily ligand-receptor network (Vilar et al., 2006). This work indicates that the key quantity that determines the qualitative behavior of the pathway is the ratio of the constitutive to the ligand-induced rate of receptor degradation (CIR, constitutive-to-induced degradation ratio). Low CIR causes a transient increase of signal activity that returns to pre-stimulus levels. On the contrast, high CIR produces a permanently elevated level of signal activity. The

concept of CIR refers to the rates of two degradations process rather than the simple expression of the degrading rate constants. This conclusion is affirmed by our analysis of the balance between clathrin dependent endocytosis versus non-clathrin mediated endocytosis. Our model shows the TGF-β signal response is regulated by the ratio of clathrin to non-clathrin endocytosis rate. On the other hand, the concise model doesn't distinguish the receptors in early endosomes and in caveolar lipid-raft. For ligand-induced degradation of receptors, the concise model simply regards the processes of the non-clathrin dependent internalization, recycling and the degradation of the receptors in caveolar lipid-raft as one-step of ligand induced receptor degradation at cell surface. Therefore, the control role of CIR ratio proposed in the concise model is consistent with regulation role of the clathrin to non-clathrin endocytosis rate ratio in the comprehensive model of this work.

Chapter 5 Conclusions and General Discussions

The understanding of signaling pathways requires both much experimental work and theoretical analysis. With the development of mathematical models and the relative software packages, we have improved our understanding of receptor trafficking network and TGF-β signaling pathway. This chapter briefly summarizes the results from this thesis, points out some open questions in modeling and kinetic analysis of cellular signaling pathways and addresses some thoughts on how to solve some of these problems.

5.1 Summary

The first step in most of cellular signaling pathways is processing the external signal (ligand) by the interaction between the ligand and the corresponding receptor at cell surface. The activated receptors initiate the signal by interacting with the downstream components of the signaling pathway and transmit the external signal into intracellular response. Some receptors can shuttle between cell surface and cytoplasm by internalization and recycling, which is termed as receptor trafficking. In order to have a simple mathematical analysis, traditional mathematical models usually ignored the extracellular medium compartment and assume that the ligand concentration in the medium is constant or follows a decaying function over time after cells are exposed to the external signal. We have systematically analyzed the dynamical behavior of the ligand and signal responses in different cell density scenarios for 8 possible receptor trafficking networks. The analyses indicate that the traditional assumption is only valid in some specific cases and should be reconsidered for different system. The work also found that receptor activity is potentially controlled by the ratio between ligand number and surface receptor number per cell.

The selection of parameter values for the mathematical models is crucial for reproducing the experimental data and making predictions from the models. In order to speed up the parameter estimation work for modeling of signaling pathways, we developed a systems biology markup language based parameter estimation tool, SBML-PET. This tool is able to perform parameter estimation by fitting a variety of experimental data and is applicable for different models.

Based on the work of receptor trafficking network and the development of SBML-PET, we established a comprehensive model for the TGF-β signaling pathway. The model considered the TGF-β receptor trafficking, Smad dependent downstream signal events and the I-Smad negative feedback regulation. The constraints-based modeling method is applied to the modeling of this pathway by transforming the qualitative experimental data to some constraints in the SBML-PET tool. The model is able to reproduce the quantitative analysis of TGF-β signaling pathway. Further kinetic analysis of the model indicates that signal response to TGF-β is regulated by the balance between clathrin dependent endocytosis and non-clathrin mediated endocytosis.

5.2 General Discussions

The receptor trafficking in cellular signaling pathways is a complicated process. For particular receptor trafficking pathways, it is important to take other complexities into account. These complexities include the cell size and shape (Meyers et al., 2006), nonlinear kinetics of receptor activation (receptor homodimerization and heterodimerization), particular receptor endocytosis dynamics, specific ligand-receptor interaction (Shankaran et al., 2007) and the potential crosstalk of the same receptor in different pathways. The activity of receptor is found to be regulated by the feedbacks of the downstream signal response (Ogunjimi et al., 2005). Therefore, it is necessary to consider these complexities for the study of the specific signal response for a particular pathway.

Parameter estimation is essentially an optimization process, which has no perfect solution for different problems. Currently, SBML-PET employs the stochastic ranking evolutionary strategy algorithm (SRES) for the optimization process. Although, SRES algorithm has some good performance for some tested biochemical models, it might perform not well for some cases. It would be better to make SBML-PET have some other global optimization algorithms, for example differential evolution. Then, the users have more options to test which algorithm is better for their specific problems.

A few years ago, the field of systems biology was trying hard to obtain some quantitative data and enable the mathematical model to reproduce the experimental data. With the development of parameter estimation software packages, it becomes easier to fit the model simulation to the experimental data now. However, over-fitting problem also appears due to the complexity of the model and the correlation of the parameters. Fitting the experimental

data well doesn't mean that the mathematical model has correctly described the corresponding system. In some cases, there are many possible parameter sets for the model to reproduce the available experimental data. The over-fitting problem might lead to some unwarranted conclusions because the different parameter sets might similarly fit the data well, but result in different predictions. Therefore, it is important to validate and calibrate the model with additional experimental data, for example, the perturbation of the system.

5.3 Future Work

Much work remains to be done for the mathematical modeling and kinetic analysis of cellular signaling pathways. Biological systems are more complicated than the simplified mathematical models and the availability of the experimental data are limited for identify the parameters in the models. Therefore, the uncertainty of the biological system and the identifiability of the parameters in the model limit our confidence of the models and their predictions. In the future, we need establish a close collaboration with the experimentalist to obtain more experimental data to calibrate the model. On the other hand, the analysis of multiple candidate models under different experimental conditions might give us some insights about difference of the possible mechanisms from these models. For example, we can perturb the models for the target signaling pathways by changing the signal input, over expressing, knocking down or knocking out of a certain species in the model which is feasible for experimental analysis (Santos et al., 2007). Since experimental work is time-consuming, *in silico* kinetic analysis of the models is useful for the optimal design of new experiments, which can discriminate alternative hypotheses underlying different models (Maiwald et al., 2007). With the combination of the quantitative analysis of mathematical models and the experimental analysis, we would be able to broaden our understanding of cellular signaling pathways.

References

Aldridge, B.B., Burke, J.M., Lauffenburger, D.A., and Sorger, P.K. (2006). Physicochemical modelling of cell signalling pathways. Nat Cell Biol *8*, 1195-1203.

Ashe, H.L., and Briscoe, J. (2006). The interpretation of morphogen gradients. Development *133*, 385-394.

Bakin, A.V., Tomlinson, A.K., Bhowmick, N.A., Moses, H.L., and Arteaga, C.L. (2000). Phosphatidylinositol 3-kinase function is required for transforming growth factor beta-mediated epithelial to mesenchymal transition and cell migration. J Biol Chem *275*, 36803-36810.

Bentele, M., Lavrik, I., Ulrich, M., Stosser, S., Heermann, D.W., Kalthoff, H., Krammer, P.H., and Eils, R. (2004). Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis. J Cell Biol *166*, 839-851.

Berkers, J.A., van Bergen en Henegouwen, P.M., and Boonstra, J. (1991). Three classes of epidermal growth factor receptors on HeLa cells. J Biol Chem *266*, 922-927.

Beyer, H.-G., and Schwefel, H.-P. (2002). Evolution strategies - A comprehensive introduction. Natural Computing *1*, 3-52.

Bhartiya, S., Rawool, S., and Venkatesh, K.V. (2003). Dynamic model of Escherichia coli tryptophan operon shows an optimal structural design. Eur J Biochem *270*, 2644-2651.

Blower, S.M., and Dowlatabadi, H. (1994). Sensitivity and Uncertainty Analysis of Complex-Models of Disease Transmission - an Hiv Model, as an Example. International Statistical Review *62*, 229-243.

Boogerd, F.C., Bruggeman, F.J., Hofmeyr, J.-H.S., and Westerhoff, H.V. (2007). Systems Biology: Philosophical Foundations, 1st edn (Amsterdam, Elsevier).

Bruggeman, F.J., and Westerhoff, H.V. (2007). The nature of systems biology. Trends Microbiol *15*, 45-50.

Burke, P.M., and Wiley, H.S. (1999). Human mammary epithelial cells rapidly exchange empty EGFR between surface and intracellular pools. J Cell Physiol *180*, 448-460.

Childress, J.L., Acar, M., Tao, C., and Halder, G. (2006). Lethal giant discs, a novel C2-domain protein, restricts notch activation during endocytosis. Curr Biol *16*, 2228-2233.

Cho, K.H., Shin, S.Y., Kolch, W., and Wolkenhauer, O. (2003). Experimental design in systems biology based on parameter sensitivity analysis using a Monte Carlo method: a case study for the TNF alpha-mediated NF-kappaB signal transduction pathway. SIMULATION *79*, 726-739.

Clarke, D.C., Betterton, M.D., and Liu, X. (2006). Systems theory of Smad signalling. Syst Biol (Stevenage) *153*, 412-424.

Clayton, A.H., Walker, F., Orchard, S.G., Henderson, C., Fuchs, D., Rothacker, J., Nice, E.C., and Burgess, A.W. (2005). Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-A multidimensional microscopy analysis. J Biol Chem *280*, 30392-30399.

Conlon, I., and Raff, M. (2003). Differences in the way a mammalian cell and yeast cells coordinate cell growth and cell-cycle progression. J Biol *2*, 7.

DeWitt, A.E., Dong, J.Y., Wiley, H.S., and Lauffenburger, D.A. (2001). Quantitative analysis of the EGF receptor autocrine system reveals cryptic regulation of cell response by ligand capture. J Cell Sci *114*, 2301-2313.

Dhar, P.K., Meng, T.C., Somani, S., Ye, L., Sakharkar, K., Krishnan, A., Ridwan, A.B., Wah, S.H., Chitre, M., and Hao, Z. (2005). Grid cellware: the first grid-enabled tool for modelling and simulating cellular processes. Bioinformatics *21*, 1284-1287.

Di Guglielmo, G.M., Le Roy, C., Goodfellow, A.F., and Wrana, J.L. (2003). Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. Nat Cell Biol *5*, 410-421.

Di Ventura, B., Lemerle, C., Michalodimitrakis, K., and Serrano, L. (2006). From in vivo to in silico biology and back. Nature *443*, 527-533.

Felder, S., LaVin, J., Ullrich, A., and Schlessinger, J. (1992). Kinetics of binding, endocytosis, and recycling of EGF receptor mutants. J Cell Biol *117*, 203-212.

Feng, X.H., and Derynck, R. (2005). Specificity and versatility in tgf-beta signaling through Smads. Annu Rev Cell Dev Biol *21*, 659-693.

Feng, X.J., Hooshangi, S., Chen, D., Li, G., Weiss, R., and Rabitz, H. (2004). Optimizing genetic circuits by global sensitivity analysis. Biophys J *87*, 2195-2202.

Finney, A., and Hucka, M. (2003). Systems biology markup language: Level 2 and beyond. Biochem Soc Trans *31*, 1472-1473.

Fischer, J.A., Eun, S.H., and Doolan, B.T. (2006). Endocytosis, endosome trafficking, and the regulation of Drosophila development. Annu Rev Cell Dev Biol *22*, 181-206.

French, A.R., Tadaki, D.K., Niyogi, S.K., and Lauffenburger, D.A. (1995). Intracellular trafficking of epidermal growth factor family ligands is directly influenced by the pH sensitivity of the receptor/ligand interaction. J Biol Chem *270*, 4334-4340.

Funahashi, A., Tanimura, N., Morohashi, M., and Kitano, H (2003). CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. BIOSILICO *1*, 159.

Gadkar, K.G., Gunawan, R., and Doyle, F.J., 3rd (2005). Iterative approach to model identification of biological networks. BMC Bioinformatics *6*, 155.

Goumans, M.J., Valdimarsdottir, G., Itoh, S., Rosendahl, A., Sideras, P., and ten Dijke, P. (2002). Balancing the activation state of the endothelium via two distinct TGF-beta type I receptors. Embo J *21*, 1743-1753.

Guldberg, C.M., and Waage, P. (1879). Concerning Chemical Affinity. Erdmann's Journal für Practische Chemie *127*, 69-114.

Gurdon, J.B., and Bourillot, P.Y. (2001). Morphogen gradient interpretation. Nature *413*, 797-803.

Hayes, S., Chawla, A., and Corvera, S. (2002). TGF beta receptor internalization into EEA1-enriched early endosomes: role in signaling to Smad2. J Cell Biol *158*, 1239-1249.

He, W., Dorn, D.C., Erdjument-Bromage, H., Tempst, P., Moore, M.A., and Massague, J. (2006). Hematopoiesis controlled by distinct TIF1gamma and Smad4 branches of the TGFbeta pathway. Cell *125*, 929-941.

Heinrich, R., and Schuster, S. (1996). The Regulation of Cellular Systems (Chapman and Hall).

Hill, A.V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. The Journal of Physiology *40*, iv-vii.

Hindmarsh, A.C., ed. (1983). ODEPACK, A Systematized Collection of ODE Solvers (Amsterdam, North-Holland).

Hindmarsh, A.C., Brown, P.N., Grant, K.E., Lee, S.L., Serban, R., Shumaker, D.E., and Woodward, C.S. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. Acm Transactions on Mathematical Software *31*, 363-396.

Hoffmann, A., Levchenko, A., Scott, M.L., and Baltimore, D. (2002). The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. Science *298*, 1241-1245.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U. (2006). COPASI--a COmplex PAthway SImulator. Bioinformatics *22*, 3067-3074.

Hu, D., and Yuan, J.M. (2006). Time-dependent sensitivity analysis of biological networks: coupled MAPK and PI3K signal transduction pathways. J Phys Chem A *110*, 5361-5370.

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A.*, et al.* (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics *19*, 524-531.

Incardona, J.P., Gruenberg, J., and Roelink, H. (2002). Sonic hedgehog induces the segregation of patched and smoothened in endosomes. Curr Biol *12*, 983-995.

Inman, G.J., Nicolas, F.J., and Hill, C.S. (2002). Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity. Mol Cell *10*, 283-294.

Ji, X., and Xu, Y. (2006). libSRES: a C library for stochastic ranking evolution strategy for parameter estimation. Bioinformatics *22*, 124-126.

Joyce, A.R., and Palsson, B.O. (2006). The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol *7*, 198-210.

Kavsak, P., Rasmussen, R.K., Causing, C.G., Bonni, S., Zhu, H., Thomsen, G.H., and Wrana, J.L. (2000). Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. Mol Cell *6*, 1365-1375.

Kholodenko, B.N., Demin, O.V., Moehren, G., and Hoek, J.B. (1999). Quantification of short term signaling by the epidermal growth factor receptor. J Biol Chem *274*, 30169-30181.

Kholodenko, B.N., Kiyatkin, A., Bruggeman, F.J., Sontag, E., Westerhoff, H.V., and Hoek, J.B. (2002). Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. Proc Natl Acad Sci U S A *99*, 12841-12846.

Kimura, S., Kawasaki, T., Hatakeyama, M., Naka, T., Konishi, F., and Konagaya, A. (2004). OBIYagns: a grid-based biochemical simulator with a parameter estimator. Bioinformatics *20*, 1646-1648.

Klipp, E., and Heinrich, R. (1994). Evolutionary optimization of enzyme kinetic parameters; effect of constraints. J Theor Biol *171*, 309-323.

Klipp, E., Herwig, R., Kowald, A., Wierling, C., and Lehrach, H. (2005a). Systems Biology in Practice: Concepts, Implementation and Application, 1st edn (Berlin, Wiley-VCH).

Klipp, E., Nordlander, B., Kruger, R., Gennemark, P., and Hohmann, S. (2005b). Integrative model of the response of yeast to osmotic shock. Nat Biotechnol *23*, 975-982.

Knauer, D.J., Wiley, H.S., and Cunningham, D.D. (1984). Relationship between epidermal growth factor receptor occupancy and mitogenic response. Quantitative analysis using a steady state model system. J Biol Chem *259*, 5623-5631.

Kremling, A., Fischer, S., Gadkar, K., Doyle, F.J., Sauter, T., Bullinger, E., Allgower, F., and Gilles, E.D. (2004). A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. Genome Res *14*, 1773-1785.

Le Roy, C., and Wrana, J.L. (2005). Clathrin- and non-clathrin-mediated endocytic regulation of cell signalling. Nat Rev Mol Cell Biol *6*, 112-126.

Lee, E., Salic, A., Kruger, R., Heinrich, R., and Kirschner, M.W. (2003). The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. PLoS Biol *1*, E10.

Lin, X., Duan, X., Liang, Y.Y., Su, Y., Wrighton, K.H., Long, J., Hu, M., Davis, C.M., Wang, J., Brunicardi, F.C.*, et al.* (2006). PPM1A functions as a Smad phosphatase to terminate TGFbeta signaling. Cell *125*, 915-928.

Mahdavi, A., Davey, R.E., Bhola, P., Yin, T., and Zandstra, P.W. (2007). Sensitivity Analysis of Intracellular Signaling Pathway Kinetics Predicts Targets for Stem Cell Fate Control. PLoS Comput Biol *3*, e130.

Maiwald, T., Kreutz, C., Pfeifer, A.C., Bohl, S., Klingmuller, U., and Timmer, J. (2007). Dynamic pathway modeling: feasibility analysis and optimal experimental design. Ann N Y Acad Sci *1115*, 212-220.

Massague, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev *19*, 2783-2810.

Matsubara, Y., Kikuchi, S., Sugimoto, M., and Tomita, M. (2006). Parameter estimation for stiff equations of biosystems using radial basis function networks. BMC Bioinformatics *7*, 230.

Mckay, M.D., Beckman, R.J., and Conover, W.J. (1979). Comparison of 3 Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics *21*, 239-245.

Melke, P., Jonsson, H., Pardali, E., ten Dijke, P., and Peterson, C. (2006). A rate equation approach to elucidate the kinetics and robustness of the TGF-beta pathway. Biophys J *91*, 4368-4380.

Mendes, P. (1993). GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. Comput Appl Biosci *9*, 563-571.

Mendes, P., and Kell, D. (1998). Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics *14*, 869-883.

Meyers, J., Craig, J., and Odde, D.J. (2006). Potential for control of signaling pathways via cell size and shape. Curr Biol *16*, 1685-1693.

Mitchell, H., Choudhury, A., Pagano, R.E., and Leof, E.B. (2004). Ligand-dependent and -independent transforming growth factor-beta receptor recycling regulated by clathrin-mediated endocytosis and Rab11. Mol Biol Cell *15*, 4166-4178.

Moles, C.G., Mendes, P., and Banga, J.R. (2003). Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res *13*, 2467-2474.

Moore, C.A., Milano, S.K., and Benovic, J.L. (2007). Regulation of receptor trafficking by GRKs and arrestins. Annu Rev Physiol *69*, 451-482.

Myers, A.C., Kovach, J.S., and Vuk-Pavlovic, S. (1987). Binding, internalization, and intracellular processing of protein ligands. Derivation of rate constants by computer modeling. J Biol Chem *262*, 6494-6499.

Nagashima, T., Shimodaira, H., Ide, K., Nakakuki, T., Tani, Y., Takahashi, K., Yumoto, N., and Hatakeyama, M. (2007). Quantitative transcriptional control of ErbB receptor signaling undergoes graded to biphasic response for cell differentiation. J Biol Chem *282*, 4045-4056.

Ogunjimi, A.A., Briant, D.J., Pece-Barbara, N., Le Roy, C., Di Guglielmo, G.M., Kavsak, P., Rasmussen, R.K., Seet, B.T., Sicheri, F., and Wrana, J.L. (2005). Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain. Mol Cell *19*, 297-308.

Pierreux, C.E., Nicolas, F.J., and Hill, C.S. (2000). Transforming growth factor beta-independent shuttling of Smad4 between the cytoplasm and nucleus. Mol Cell Biol *20*, 9041-9054.

Price, N.D., Papin, J.A., Schilling, C.H., and Palsson, B.O. (2003). Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol *21*, 162-169.

Runarsson, T.P., and Yao, X. (2000). Stochastic ranking for constrained evolutionary optimization. Ieee Transactions on Evolutionary Computation *4*, 284-294.

Salazar, G., and Gonzalez, A. (2002). Novel mechanism for regulation of epidermal growth factor receptor endocytosis revealed by protein kinase A inhibition. Mol Biol Cell *13*, 1677-1693.

Santos, S.D., Verveer, P.J., and Bastiaens, P.I. (2007). Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. Nat Cell Biol *9*, 324-330.

Savini, I., Duflot, S., and Avigliano, L. (2000). Dehydroascorbic acid uptake in a human keratinocyte cell line (HaCaT) is glutathione-independent. Biochem J *345 Pt 3*, 665-672.

Sawano, A., Takayama, S., Matsuda, M., and Miyawaki, A. (2002). Lateral propagation of EGF signaling after local stimulation is dependent on receptor density. Dev Cell *3*, 245-257.

Schlessinger, J. (1988). The epidermal growth factor receptor as a multifunctional allosteric protein. Biochemistry *27*, 3119-3123.

Schmierer, B., and Hill, C.S. (2005). Kinetic analysis of Smad nucleocytoplasmic shuttling reveals a mechanism for transforming growth factor beta-dependent nuclear accumulation of Smads. Mol Cell Biol *25*, 9845-9858.

Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., and Muller, G. (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol *20*, 370-375.

Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol *3*, 119.

Seto, E.S., Bellen, H.J., and Lloyd, T.E. (2002). When cell biology meets development: endocytic regulation of signaling pathways. Genes Dev *16*, 1314-1336.

Shankaran, H., Resat, H., and Wiley, H.S. (2007). Cell Surface Receptors for Signal Transduction and Ligand Transport: A Design Principles Study. PLoS Comput Biol *3*, e101.

Shi, Y., and Massague, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell *113*, 685-700.

Soderquist, A.M., and Carpenter, G. (1986). Biosynthesis and metabolic degradation of receptors for epidermal growth factor. J Membr Biol *90*, 97-105.

Starbuck, C., and Lauffenburger, D.A. (1992). Mathematical model for the effects of epidermal growth factor receptor trafficking dynamics on fibroblast proliferation responses. Biotechnol Prog *8*, 132-143.

Swameye, I., Muller, T.G., Timmer, J., Sandra, O., and Klingmuller, U. (2003). Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. Proc Natl Acad Sci U S A *100*, 1028-1033.

Tan, C.M., Brady, A.E., Nickols, H.H., Wang, Q., and Limbird, L.E. (2004). Membrane trafficking of G protein-coupled receptors. Annu Rev Pharmacol Toxicol *44*, 559-609.

Thomas, S., Mooney, P.J., Burrell, M.M., and Fell, D.A. (1997). Metabolic Control Analysis of glycolysis in tuber tissue of potato (Solanum tuberosum): explanation for the low control coefficient of phosphofructokinase over respiratory flux. Biochem J *322 (Pt 1)*, 119-127.

Tsai, K.Y., and Wang, F.S. (2005). Evolutionary optimization with data collocation for reverse engineering of biological networks. Bioinformatics *21*, 1180-1188.

Tzafriri, A.R., and Edelman, E.R. (2007). Endosomal receptor kinetics determine the stability of intracellular growth factor signalling complexes. Biochem J *402*, 537-549.

Vilar, J.M., Jansen, R., and Sander, C. (2006). Signal processing in the TGF-beta superfamily ligand-receptor network. PLoS Comput Biol *2*, e3.

Wakefield, L.M., Smith, D.M., Masui, T., Harris, C.C., and Sporn, M.B. (1987). Distribution and modulation of the cellular receptor for transforming growth factor-beta. J Cell Biol *105*, 965-975.

Watt, F.M., Jordan, P.W., and O'Neill, C.H. (1988). Cell shape controls terminal differentiation of human epidermal keratinocytes. Proc Natl Acad Sci U S A *85*, 5576-5580.

Wiley, H.S. (2003). Trafficking of the ErbB receptors and its influence on signaling. Exp Cell Res *284*, 78-88.

Wiley, H.S., and Cunningham, D.D. (1981). A steady state model for analyzing the cellular binding, internalization and degradation of polypeptide ligands. Cell *25*, 433-440.

Xu, L., Chen, Y.G., and Massague, J. (2000). The nuclear import function of Smad2 is masked by SARA and unmasked by TGFbeta-dependent phosphorylation. Nat Cell Biol *2*, 559-562.

Xu, L., Kang, Y., Col, S., and Massague, J. (2002). Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGFbeta signaling complexes in the cytoplasm and nucleus. Mol Cell *10*, 271-282.

Zak, D.E., Gonye, G.E., Schwaber, J.S., and Doyle, F.J., 3rd (2003). Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: insights from an identifiability analysis of an in silico network. Genome Res *13*, 2396-2405.

Zheng, Y., and Rundell, A. (2006). Comparative study of parameter sensitivity analyses of the TCR-activated Erk-MAPK signalling pathway. IEE Proc Syst Biol *153*, 201-211.

Zi, Z., Cho, K.H., Sung, M.H., Xia, X., Zheng, J., and Sun, Z. (2005). In silico identification of the key components and steps in IFN-gamma induced JAK-STAT signaling pathway. FEBS Lett *579*, 1101-1108.

Zi, Z., and Klipp, E. (2006). SBML-PET: a Systems Biology Markup Language-based parameter estimation tool. Bioinformatics *22*, 2704-2705.

Zi, Z., and Klipp, E. (2007a). Cellular signaling is potentially regulated by cell density in receptor trafficking networks. FEBS Lett *581*, 4589-4595.

Zi, Z., and Klipp, E. (2007b). Constraint-based modeling and kinetic analysis of the smad dependent tgf-Beta signaling pathway. PLoS ONE *2*, e936.

Appendix A: Latin Hypercube Sampling

Latin hypercube sampling (LHS) is an efficient method to sample random parameter vectors while guaranteeing that individual parameter ranges are evenly covered. It was developed to address the need of uncertainty assessment for a particular class of problems. This method is first developed by McKay, Conover in 1979 (Mckay et al., 1979).

Suppose a variable y that is a function of other variables x_1, x_2, …, x_n (usually we called them parameters in the field of systems biology). This function may be very complicated, for example, a set of ordinary differential equations for biological models. A question to be investigated is "How does y vary when the variables x vary according to some assumed joint probability distribution?" Traditionally, we can use Monte Carlo sampling to study this question. By repeatedly sampling from the assumed distribution of the parameters and evaluating y for each sample, the behavior of y can be investigated. Latin hypercube sampling is an alternative approach, which can yield more precise estimates. Generally, LHS will require fewer samples than simple Monte Carlo sampling to achieve a similar accuracy.

The procedure of Latin hypercube sampling for selecting M different values from each of N variables x_1, x_2, …, x_n can be summarized as:

(1) Divide the range of each variable into M equally probable intervals.

(2) From each interval, select a value randomly with respect to the probability density in the interval.

(3) The M values thus obtained for x_1 are paired randomly (similar as combinations) with the M values of x_2. These M pairs are combined in a random manner with the M values of x_3 to form M triplets, and so on, until the (M×N) matrix is formed. It is convenient to think of this sample as forming an (M×N) matrix of input where the *i*th row contains specific values of the N input variables to be used in the *i*th simulation.

For example, if we want to generate 10 samples for each of the three variables, X1, X2 and X3, which corresponds to a 10×3 matrix. We assume the variation ranges of these three variables are between 0 and 1. If the samples are designed by Latin hypercube sampling method, it will make sure that the sampling points of X1, X2 and X3 will be equally distributed between 0 and 1 (Figure App.1A), which means that X1, X2 and X3 will have one

and only one sample at each of the 10 intervals (0, 0.1], (0.1, 0.2], (0.2, 0.3], …, (0.8, 0.9], (0.9, 1). However, if we randomly generate the samples from a uniform distribution by the general Monte Carlo sampling, there are more than one samples falling into some intervals and some intervals might have no any samples for X1, X2 or X3 (Figure App.1B).

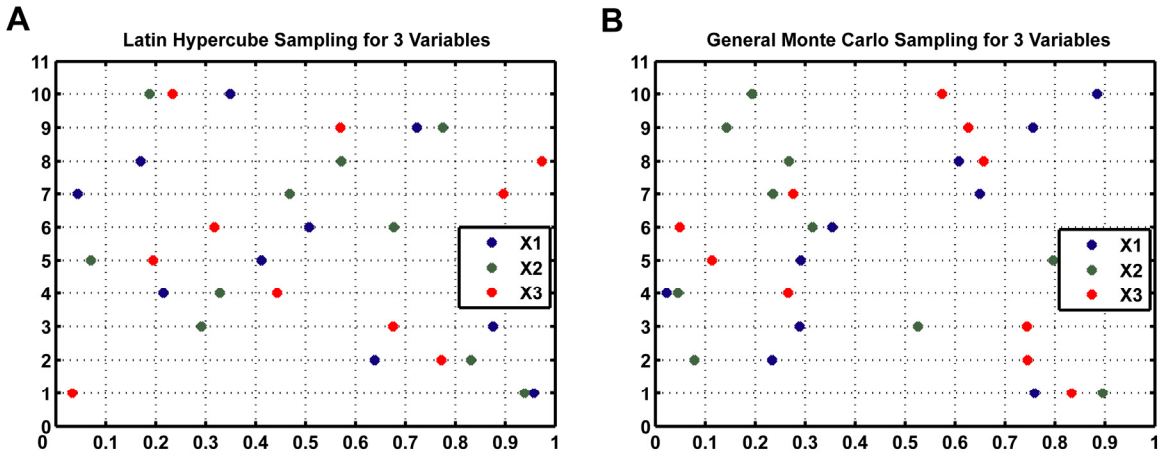

Figure App.1 10×3 samples design for 3 variables by different sampling methods. (A) Latin hypercube sampling. (B) General Monte Carlo sampling from a uniform distribution.

Appendix B: Stochastic Ranking Evolution Strategy (SRES) Algorithm

The Stochastic Ranking Evolution Strategy (SRES) algorithm is proposed by Runarsson and Yao in 2000 and the description of SRES here is written based on their work (Runarsson and Yao, 2000). SRES is a modified evolutionary optimization method using stochastic ranking technique to handle the constraints. The central optimization process employs the classic (μ, λ)-ES evolution strategy algorithm, in which the selection is taken from the λ offspring only, whereas their μ parents are ignored even the fitness of the parents are better than that of the new generation (Beyer and Schwefel, 2002).

The parameter estimation for the cellular signaling pathway models is actually a optimization issue, which can summarized as minimizing the objective function:

$$\text{minimize } f(X), \quad X = (x_1, ..., x_n) \in R^n \tag{App.1}$$

where *f(X)* is the objective function, which is usually the (weighted) least square error between the model simulation and the experimental data (see equation 1.9 in Chapter 1); $X = (x_1, ..., x_n) \in R^n$ is the parameter set defines the search space which is a *n*-dimensional space bounded by the range of the parameters:

$$x_i^{\min} \le x_i \le x_i^{\max}, \quad i \in \{1, ..., n\} \tag{App.2}$$

and the feasible region *F*, which is defined by the constraints

$$F = \{X \in R^n \mid g_j(X) \le 0; \quad j \in \{1, ..., m\}\} \tag{App.3}$$

where $g_j(X) \le 0, j \in \{1, ..., m\}$ are the constraints.

Traditional methods deal with the constrained optimization problem by adding a penalty term into the objective function to penalize constraint violation. Then, the constrained optimization problem is transformed into an unconstrained one with new objective function *ψ(X)*.

$$\psi(X) = f(X) + r_k p(g_j(X)); \quad j = 1, ..., m \tag{App.4}$$

where $p \geq 0$ is a real valued function which imposes a 'penalty' regulated by the penalty coefficient $r_k \geq 0$, k is the generation counter.

For two individuals i, i+1 in the λ individuals, let us compare their transformed objective function in ranked order according to (App.4).

$$f_i + r_k p_i < f_{i+1} + r_k p_{i+1} \quad i \in \{1, ..., \lambda - 1\} \tag{App.5}$$

Here the notation $f_i = f(X_i)$ and $p_i = p(g_j(X_i))$ are used for convenience. The inequality (App. 5) can be reformatted as the following for $p_i \neq p_{i+1}$

$$(r_k - \hat{r})(p_i - p_{i+1}) < 0 \quad \text{for } p_i \neq p_{i+1} \tag{App.6}$$

$$\hat{r} = \frac{f_{i+1} - f_i}{p_i - p_{i+1}}$$

where $\hat{r}$ is the critical penalty coefficient.

For the choice of r_k, there are three different cases which may make the inequality (App.5) to be valid:

(1) $p_i \leq p_{i+1}$ and $f_i \geq f_{i+1}$, the comparison is said to be *dominated by penalty function* and $r_k > \hat{r} > 0$ because the penalty function p plays the dominate role in evaluating the inequality. In this case, if we set $r_k < \hat{r}$, then r_k is too small so that the penalty function can not influence the ranking of the individuals, the inequality (App.5) will not be valid. Then it is called *under-penalization*.

(2) $p_i \geq p_{i+1}$ and $f_i \leq f_{i+1}$, the comparison is said to be *dominated by objective function* and $\hat{r} > r_k > 0$ because the objective function f plays the dominate role in evaluating the inequality. In this case, if $r_k > \hat{r}$, r_k is too large that the impact of the objective function can be ignored and the inequality (App.5) will not be valid. Then it is called *over-penalization*.

(3) $p_i \leq p_{i+1}$ and $f_i \leq f_{i+1}$, the comparison is said to be *non-dominated* and $\hat{r} < 0$ because the inequality (App.5 or App.6) is always valid in this case.

Therefore, not all possible r_k can affect the ranking of the individuals. The value of r_k should be within a certain range, not too large and not too small ($\hat{r}_{\min} < r_k < \hat{r}_{\max}$) and such an optimal r_k is hard to determine in practice. SRES algorithm introduces a different approach to balance the dominance of the objective and penalty function. Given any pair of two adjacent individuals, the probability of comparing them according to the objective function f is 1 if both individuals are feasible ($p_i = 0$ and $p_{i+1} = 0$), otherwise, it is P_f, which is a probability of using only the objective function f for comparison in ranking the infeasible regions of the search space ($p_i > 0$ or $p_{i+1} > 0$). The ranking of the λ individuals is achieved by a bubble-sort-like procedure, which is a stochastic ranking technique and shown in Figure App.2.

```
for i=1:N
    for j=1:λ-1
        random number u from U(0,1)
        if (p_i = 0 and p_{i+1} = 0) or u<P_f
            if  f(X_j) > f(X_{j+1})
                swap(X_j, X_{j+1})
            else
                if  p(X_j) > p(X_{j+1})
                    swap(X_j, X_{j+1})
                end
            end
        end
    end
end
```

Figure App.2 Stochastic ranking of the λ individuals in SRES algorithm. U(0,1) is a uniform random number generator and N is number of sweeps going through the whole population. Adapted from Fig. 1 in Runarsson and Yao's work (Runarsson and Yao, 2000).

The value of the P_f can control the contribution of the penalty function. The ranking is an *over-penalization* and *under-penalization* for $P_f = 0$ and $P_f = 1$, respectively. $P_f = 0.45$ is a good choice for most of the problems. However, the optimal of P_f value is slightly varied for different optimization problems.

Acknowledgments

I would like to thank the support and help from the whole team of Klipp's group, especially Dr. Edda Klipp for her guidance throughout the past three years of my research. The members of the group have done me many favors for my research and daily life.

I am indebted to Prof. Dr. Martin Vingron for granting me the scholarship from the program of International Max Planck Research School for Computational Biology & Scientific Computing (IMPRS). I also would like to acknowledge Hannes Luz for his assistance during my study in Berlin.

In addition, I appreciate the help from our collaborator Aristidis Moustakas for his suggestion and discussion on kinetic analysis of the TGF-β signaling pathway work.

From the bottom of my heart, I want to express my gratitude to my wife, Difan Deng. My life in Berlin would be less comfortable without her accompany in Berlin. Her love, care and support are immeasurable.

Curriculum vitae

PERSONAL INFORMATION	
Name	Zhike Zi
Work Address	Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195, Berlin, Germany
Gender	Male
Date of Birth	December 27, 1979
Place of Birth	HUNAN, P. R. CHINA
Nationality	P. R. CHINA

EDUCATION	
1998 – 2002	BSc. in Biotechnology at College of Chemistry and Life Science, South-central University for Nationalities, Wuhan, China.
2002 – 2005	MSc. in Bioinformatics at Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing, China
2005 – 2008	PhD Candidate, Max Planck Institute for Molecular Genetics, Berlin, Germany

RESEARCH EXPERIENCE	
2002 – 2005	MSc. thesis project about global sensitivity analysis and robustness analysis of JAK-STAT signaling pathway (Tsinghua University, Beijing, China)
July, 2006	DAAD exchange program about tools development for systems biology (Virginia Bioinformatics Institute Center, Virginia Polytechnic Institute and State University, USA)
2005 - 2008	PhD thesis project (Berlin, Germany)

LIST OF PUBLICATIONS

Papers in submission and preparation

Zhike Zi, Yanan Zheng, Ann E Rundell and Edda Klipp. (2008). SBML-SAT: a Systems Biology Markup Language (SBML) based sensitivity analysis tool. *BMC Bioinformatics* (in revision).

Miquel Àngel Adrover*, Zhike Zi*, Jörg Schaber, Josep Clotet, Edda Klipp and Francesc Posas. (2008). A quantitative approach reveals a distinct temporal role for cyclins and Sic1 to control G1-S progression upon stress. (* contributed equally, in preparation)

Publications during PhD study

Zhike Zi, Edda Klipp. (2007). Constraint-based modeling and kinetic analysis of the smad dependent TGF-β signaling pathway. *PLoS ONE*, **2** (9):e936.

Zhike Zi, Edda Klipp. (2007). Cellular signaling is potentially regulated by cell density in receptor trafficking networks. *FEBS Letters*, 581 (24):4589-95.

Zhike, Zi, Edda Klipp. (2007). Steady state analysis of signal response in receptor trafficking networks. *Genome Informatics*, 18(1), 100-108.

Zhike Zi, Edda Klipp. (2006). SBML-PET: a Systems Biology Markup Language-based parameter estimation tool. *Bioinformatics*, 22, p. 2704-2705.

Publications before PhD study

Zhike Zi, Kwang-Hyun Cho, Myong-Hee Sung, Xuefeng Xia, Jiashun Zheng and Zhirong Sun. (2005). In silico identification of the key components and steps in IFN-γ induced JAK-STAT signaling pathway. *FEBS Letters*, 579(5):1101-1108.

Zhike Zi, Zhirong Sun. (2005). Robustness Analysis of IFN-γ induced JAK-STAT Signaling Pathway. *Journal of Computer Science & Technology*, 20(4):491-495.

Statement

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe.

Zhike Zi

Berlin, 21.04.2008